SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I3P115

Volume 13 Issue 3, 185-192, March 2026

Original Article

Handcrafted-Guided, Bias-Aware Cross-Attention for Token-Level Pre-Pooling Fusion in Breast Histopathology

Pattupogula Subramanyam[1], Gurumurthy Hari Krishnan[2]

[1]*Department of Electronics and Communication Engineering, Mohan Babu University, Tirupati, India.*
[2]*Department of Electrical and Electronics Engineering, Mohan Babu University, Tirupati, India.*

[2]*Corresponding Author : haris_eee@yahoo.com*

Received: 11 December 2025 Revised: 14 January 2026 Accepted: 17 February 2026 Published: 23 March 2026

Abstract *- Histopathological evaluations of breast cancer are the most accurate form of evaluation; However, the current state of automated computer-assisted diagnostics does not perform as well when there is variability between different colour stains or due to changes in magnification during image capturing. Most current deep learning methods require convolutional or transformer-based representations, with feature fusion performed after spatial pooling of images, which hinders the colour-morphology interactions needed to classify normal tissue versus abnormal tissue correctly. To counteract these limitations, A novel Handcrafted Computing (HCC) module was developed, which fuses (pre-pooled) tokens created using HCC-generated compact handcrafted colour statistic features and deep CNN-generated spatial feature tokens. The new design enables DCOG to model persistent colour-region affinities using a learnable attention bias prior, this enables handcrafted generated tokens to work as first-class queries for spatial tokens while remaining lightweight and backbone agnostic; The proposed framework has been validated on the publicly available BreaKHis breast cancer histopathological dataset at each of the defined four magnification factors, 40×; 100×; 200×; 400× using patient-level splits to prevent data leakage between training and test sample sets and the performance was assessed using the metrics of Accuracy, Macro F1 Score, Macro AUC and Cohen's Kappa to measure both discriminative ability as well as agreement beyond chance level. Experiments have shown that the proposed HCA module reaches an accuracy of 96.87%, Macro-F1 of 96.87%, Macro-AUC of 0.9956, and a Cohen's κ of 0.9373, thereby proving superiority over traditional cross-attention without bias as well as self-attention methods. It is a remarkable achievement of the proposed technique as it has resulted in a considerably reduced number of misclassifications of critical benign lesions as malignant, while still having a high sensitivity for the malignant class.*

Keywords *- BreakHis, Breast Cancer Histopathology, Cross-Attention Fusion, Learnable Priors, Token-Level Pre-Pooling.*

1. Introduction

Histopathology of breast cancer has remained a gold standard diagnostic and treatment option for breast cancer. Variations in outcomes due to the use of specific staining methods, imaging, or the magnification settings (40x – 400x) during imaging are still present with the current histopathology process. In addition, while most of the current variable learning techniques produce high accuracy within their initial learning model's distribution, they cannot handle image variations based on their colouring and their imaging settings. Therefore, although pathologists may approve of a model based solely on the model's accuracy, it must also meet the robustness criteria required by pathologists in order to approve the model. About the use of the best performing baseline models, the late fusing process that is used to generate a classification with the use of spatial pooling can eliminate some very important colour-to-colour interactions (e.g., most importantly, the intensity of eosin, haematoxylin, and the geometry of the microstructure).

The objectives of this study are to design and investigate the effectiveness of a lightweight module that performs the pre-pooling fusion operation between the handcrafted colour statistics and the spatial feature, in a way that the visually separable differences in the feature maps are maintained in the decision. For that, this light weight module should be able to capture the color statistics in a minimal way using a few handcrafted tokens, enable the process of token choice using cross-attention from the handcrafted tokens to the spatial tokens, implement the bias prior in attention to represent the color mappings in the regions, and accomplish the above in a light weight, modular, and framework-agnostic way regardless of different CNN models. The evaluation of the experiment conducted in the BreakHis database will include

the measures of accuracy, macro-F1, macro-AUC, Cohen's-κ, as well as the comparison of the confusion matrices together with the curves of the ROC, to better examine the phenomena related to asymmetry in classification accuracy and the respective rates, which have more medical significance.

1.1. Literature Review

Classification of histopathological images of Breast Cancer can prove to be very essential for diagnosis, and the emergence of benchmark datasets like BreaKHis allows for the comparison of automatic methods for classification across various levels of magnification and types of tumours [1]. Initial research using deep learning proved the use of Convolutional Neural Networks (CNNs) in identifying key morphological differences in histopathological images using transfer learning models for training [2, 3]. Further advancements using CNN models improved the classification results, thereby proving the efficacy of CNN models for histopathological classification tasks [4, 5]. However, CNN-only models are highly sensitive to stain variability resulting from differences in tissue preparation, staining protocols, and scanning devices. With that in mind, stain normalisation and colour adaptation were proposed to reduce domain shift and improve generalisation [6, 7]. Fine-tuning strategies combined with normalisation-aware training pipelines further stabilised model performance across heterogeneous datasets [8]. Most of the presented approaches achieve stain robustness not through preprocessing but intrinsically by learning.

In an effort to increase the discriminative ability, attention mechanisms were introduced in CNN frameworks to highlight the diagnostically relevant spatial and channel-wise features. Attention-based CNN models further improved the localisation of salient tissue regions and enabled superior classification results [9, 10]. Recently, transformer architectures have significantly developed this direction due to the possibility of explicit modelling of long-range dependencies, enabling better aggregation of spatial evidence on histopathology images [11, 12].

Some works have utilised domain priors for better interpretability and robustness, specifically by incorporating nuclei-level constraints into attention-driven models [13]. Multi-branch networks and spectral attention networks were introduced for capturing complementary texture and frequency information related to glandular and stromal components [14, 15]. Learning with ensembles was also introduced for enhancing robustness through variance reduction among the breast cancer subtypes [16]. More recent work focused on hybrid models that combined CNNs and transformer architectures better to utilise local morphological features and global contextual cues. Multi-feature fusion models showed better result improvement by maintaining diverse features while learning [17, 18]. Variants of the transformer model specifically adapted to histology applications, such as Swin Transformer and Vision Transformers, have shown competitive performance with proper learning schemes [19].

Dual attention and weakly-supervised MIL models alleviated the dependency on annotations and scaled to more images, demonstrating stronger instance-level aggregation performance over standard pooling strategies [20-22]. Learnt pre-training tasks and stain-wise optimisation techniques helped to further improve, especially varying colour robustness [23]. Recent work explicitly characterised stain distributions through joint stain augmentation and normalisation to possess stronger cross-domain generalisation [24]. Nuclear-physis transformer frameworks further outperformed the discriminative strength through strategies favouring biologically important cellular patterns [25], although wavelet-based transformers showed enhancement through multi-resolution focus [26].

Transparent neighbourhood attention strategies further alleviated computational overhead without harming the ability to capture strong global-local relationships [27]. In spite of the similarities among the family of attention-based strategies, MIL-based strategies, and the family of stain-based strategies, the first is the lack of processes on the interaction between colour and spatial information or patterns that remain post-spatial pooling; precisely these aspects presumably contain subtle information regarding diagnosis. Second is the use of handmade colour statistical information. While cheap computationally and interpretable easily, handmade information is not likely to be placed first or first-class among tokens; more likely to fall under the preprocessing bucket. Thirdly, the focus is less likely to profit from the learnable bias field since the region colour correspondences are invariant on slides as well as scales.

1.2. Problem Statement

Problem: The problem addressed in the proposed research work is the development of the fusion layer in the compact plugin solution in the prePooling process and the crafted colour tokens in the CNN spatial tokens at the trainable attention-bias before boosting the resistance in the BreakHis dataset regarding the magnification factor without adding costs in the computational process. In order to accomplish this, the Handcrafted-Guided Cross-Attention (HCA) module was implemented. In this module, a bias vector is incorporated into the cross-attention calculation to facilitate a balanced relationship among colours and regions. The fusion process is conducted before spatial pooling.

Instead, the output is passed to a residual feed-forward network and a small network head. In this design, a parameter-effective module that can easily stack upon a successful backbone network was proposed. The cross-attention network with no bias vector and the self-attention mechanism are used to suppress the clinically important transition from benign to malignant, while still focusing on attribution-based interpretability.

1.3. Dataset

The experiments use the BreakHis dataset with eight histological subclasses and four magnification factors. In each modelling task, patient-level stratification is used to prevent leakage, along with stratification regarding sampling factors and the use of magnification factors. Statistics regarding normalisations are calculated from the training sets. Performed metrics include Accuracy, Macro F-1, Macro AUC, Cohen's κ, and ROC curves to analyze the confusion matrices. Data utilised is licensed from the source and is obtained from the mirror sites of the main source to enable easy access.

2. Materials and Methods

2.1. Data, Splits, and Preprocessing

Data, Splits, and For the case of BreakHis, the database comprises RGB histopathological images with a magnification level ranging from 40x, 100x, 200x, to 400x, with eight classes in total (four for the benign and four for the malignant cases). To avoid any potential information leakage, patient-level stratification is performed in all the experiments that follow. That is, no images are included in the training, validation, or testing sets if the images are from the same patient.

To facilitate this, the magnifications are kept equal in the training and validation/testing splits. When resizing the images to have a resolution equal to the resolution of the encoder, the aspect ratios are maintained through centre cropping, if needed. The parameters used in intensity normalisation are obtained through training, which are then utilised for the computation of the normalisation constants in the validation set and the test set.

To avoid any distributional information, rotation, vertical/horizontal flipping, and simple modifications in the colour space occurring within the viable limits histologically are performed only in the training set. For balancing the classification set, weightage in terms of the inverse frequency of the classes is performed during the loss calculation.

2.2. Backbone Feature Extraction and Tokenisation

A convolutional backbone (e.g., ResNet-50) produces a dense feature map $F \in \mathbb{R}^{h \times w \times d}$ from each preprocessed image. The map is flattened along spatial dimensions to obtain $C = h \cdot w$ spatial tokens $X_c \in \mathbb{R}^{C \times d}$, where each token retains local morphology and texture associated with a receptive field. In parallel, the input image is summarised by a compact RGB histogram with a small, fixed number of bins per channel; this vector is linearly projected to the model dimension to yield H handcrafted tokens. $X_h \in \mathbb{R}^{H \times d}$.

This asymmetric tokenisation, with a few handcrafted tokens and many spatial tokens, keeps the handcrafted branch extremely lightweight while allowing it to guide selection over rich spatial descriptors produced by the backbone.

2.3. Directional Projections for Cross-Attention

Directional design was adopted, in which handcrafted tokens act as queries and spatial tokens act as keys/values. With learned projections $W_Q, W_K, W_V \in \mathbb{R}^{d \times d}$, the projections are computed as

$$Q = X_h W_Q, K = X_c W_K, V = X_c W_V. \quad (1)$$

Equation (1) realizes the intended asymmetry colour statistics query morphology so that stain/colour priors explicitly drive which spatial regions are emphasised, following standard attention formulations while preserving the role assignments specified to the module.

2.4. Handcrafted-Guided, Bias-Aware Cross-Attention (Pre-Pooling Fusion)

The core fusion layer is a multi-head cross-attention that introduces a learnable bias on the token affinity matrix before normalisation. Let $B \in \mathbb{R}^{1 \times \text{heads} \times H \times C}$ denote a trainable bias tensor broadcast across the batch. The attention output is computed as

$$\text{Attn}(X_h, X_c) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}} + B\right)V \quad (2)$$

Equation (2) carries out token fusion before the spatial pooling at the token level to capture the colour-morphology interactions of the region that are otherwise averaged out in the late fusion process of Equation (2). The bias tensor B captures the colour region affinities at the hand-to-spatial pair level and is learned from the data in an end-to-end setup (initialised close to zero). Adding the bias tensor B to the softmax acts as an attention weight that is stable in query-to-region selection in the slides and scales of the object of interest, without adding extra overhead cost compared to the deep object detector models.

2.5. Post-Attention Integration and Classification Head

The model has a standard optimisation schedule (Adam with weight decay and cosine/stepwise learning rate schedule). In addition, the last stages of the learned base are left unfrozen, and the early stages are frozen during training. Early stopping is employed to monitor the macro validation F1. Class weights are estimated from the training distribution to handle the issue of imbalance. Test-time augmentation is not performed. The parameters of the normalisation modules are learned from the training split. Only the training augmentation is performed, and the case of similarities is checked inside splits to avoid memorisation. The parameter count, FLOPs, and the epoch timings are logged to measure the overhead introduced by the new layer.

2.6. Training, Optimisation, and Reproducibility

Optimisation is conducted according to a standard routine (e.g., Adam regularised by weight-decay and a cosine /

stepwise learning rate schedule). To control for balance and flexibility, the last stages in the backbone networks are unfrozen and trained, but the earlier convolutional layers are frozen. Early stopping is implemented to validate the macro F1 score.

The training distribution-derived class weighting helps handle the issue of class imbalance. There is also no use of test-time augmentation. All normalisation parameters and layer-wise counts for near-duplicate checks and FLOPs computations would only involve the training split.

2.7. Evaluation Protocol

Accuracy, macro-F1, macro-AUC (one-vs-rest), and Cohen's κ are used to gauge performance, with accompanying confusion matrices and ROC plots. Scale-stratified analyses for M-fold overall means are available to unwrap scale-dependent phenomena. Figure 1 indicates the methodology implemented in practice in this pipeline: On the left is the process of generating spatial tokens from spatial images by the backbone network. On the right is interaction with colour priors to produce histogram projections to handcrafted spatial tokens.

In the middle is bias across attention by handcrafted queries over spatial keys/values to a bias-aware attention module described by (2). On the right is the processing of residuals FFNs, pooling twice to generate final classification inputs by classification-head. The detail in this diagram concerning pre-pool fusion, query to key discrepancies with bias learning, is paralleled by definitions of X_h, X_c, Q, K, V, and B in Equations (1), (2).

2.8. Baselines, Ablations, and Intended Advantages

The proposed attention layer is then benchmarked against two internal baselines with the same split and training configuration: (i) the cross-attention layer without the bias term B, and (ii) the stacked self-attention module over the spatial tokens. The ablation includes changing the size of the designed tokens H, deleting or fixing the bias term B, and adjusting the pool operation. By design, the pre-pooling fusion remains the evidence in the local context, the hard-coded queries introduce the priors about the stains in the token selection, and the bias term regularises the mappings for the colour regions.

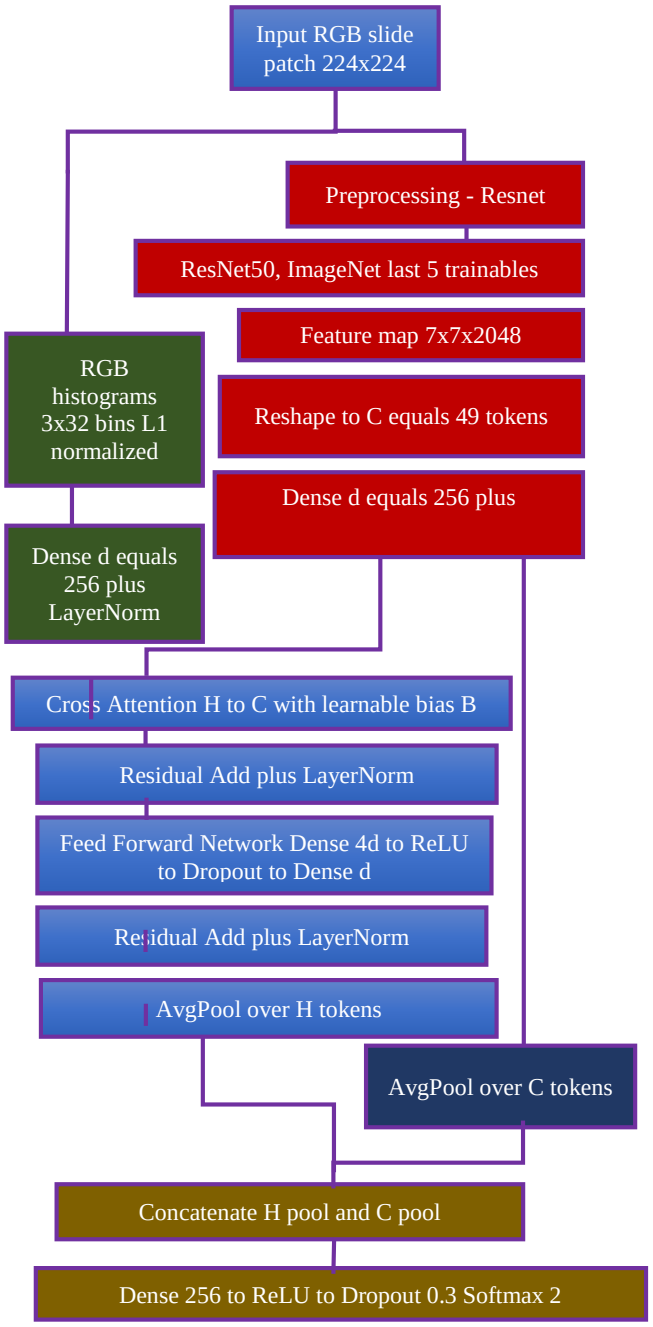

Fig. 1 The proposed handcrafted-guided, bias-aware cross-attention (pre-pooling fusion)

3. Results and Discussion

Two preliminary attention variants against the proposed Handcrafted-aware Cross-Attention (HCA) fusion on the held-out BreakHis test set (n = 447; Benign = 218, Malignant = 229). The first baseline, Cross-Attn (no bias), directly queries CNN spatial tokens with handcrafted histogram tokens but omits the learnable bias prior. The second, Self-Attn (stacked), applies attention within the concatenated token sequence without enforcing the directional handcrafted CNN design.

As highlighted in Table 1 (Cross-Attn) and Table 2 (Self-Attn), both baseline models give competing precision/recall for each class, but their respective error characteristics are quite different. The Cross-Attn baseline achieves a high precision of 0.985 for Benign, but a relatively low recall of 0.908, meaning a conservative decision boundary was introduced to resist false alarms for Benign, for 20 flips for Benign and Malignant in Figure 2: Confusion Matrix Cross-Attn (no bias).

Table 1. Cross-attn (no bias): per-class precision / recall / specificity / F1 / support

Cross-Attn (no bias) Class	Precision	Recall	Specificity	F1	Support
Benign	0.9851	0.9083	0.9869	0.9451	218
Malignant	0.9187	0.9869	0.9083	0.9516	229

Table 2. Self-attn (stacked): per-class metrics

Self-Attn (stacked) Class	Precision	Recall	Specificity	F1	Support
Benign	0.9537	0.945	0.9563	0.9493	218
Malignant	0.9481	0.9563	0.945	0.9522	229

Table 3. Proposed HCA: per-class metrics (displayed above)

Proposed HCA Class	Precision	Recall	Specificity	F1	Support
Benign	0.9595	0.9771	0.9607	0.9682	218
Malignant	0.9778	0.9607	0.9771	0.9692	229

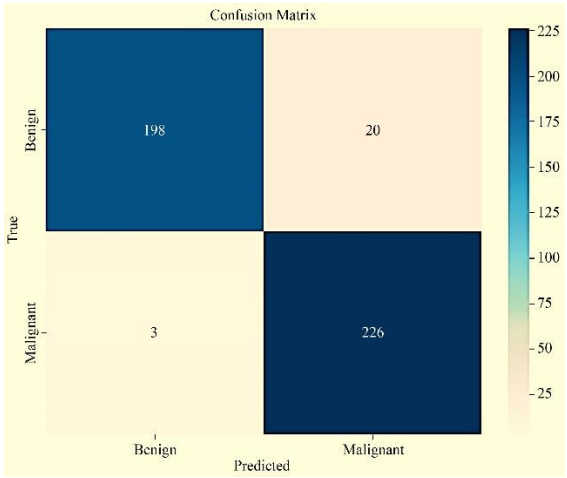

Fig. 2 Confusion matrix - cross-attn (no bias)

Conversely, Malignant sensitivity is very high (0.987), but precision is reduced (0.919) due to those same Benign flips, consistent with the overall report (Accuracy = 0.9485, Macro-F1 = 0.9483, κ = 0.8968, AUCmacro = 0.9951; Table 4).

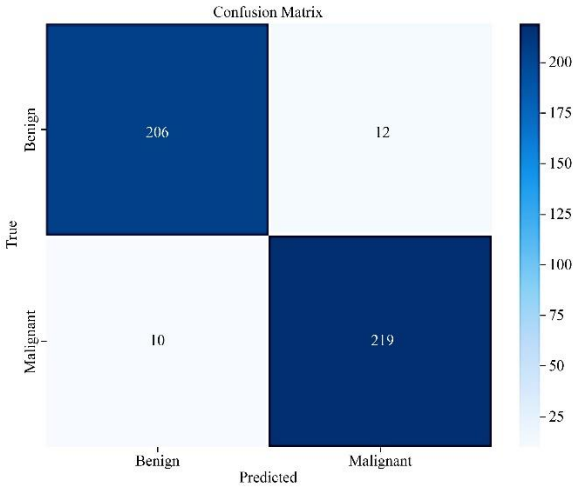

Fig. 3 Confusion matrix - self-Attn (stacked)

The Self-Attn model in Figure 3 exhibits a more symmetric error pattern (12 Benign-Malignant and 10 Malignant-Benign), reflected in Benign recall = 0.945 and Malignant recall = 0.956 with closely matched precision values (0.954 vs 0.948). This balances sensitivity across classes and yields a modest aggregate lift over Cross-Attn (Accuracy = 0.9508, Macro-F1 = 0.9507, κ = 0.9015; Table 4), although it pays a small price in separability (AUCmacro = 0.9918).

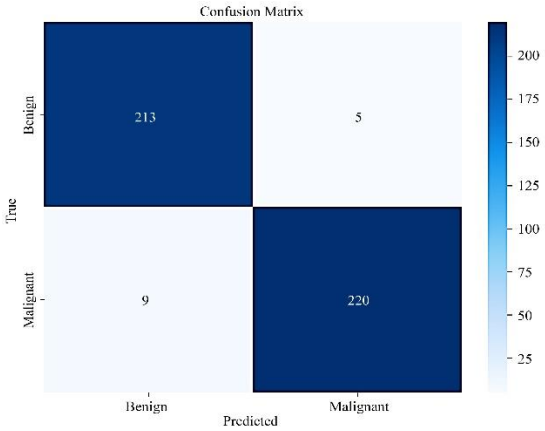

Fig. 4 Confusion matrix - proposed HCA

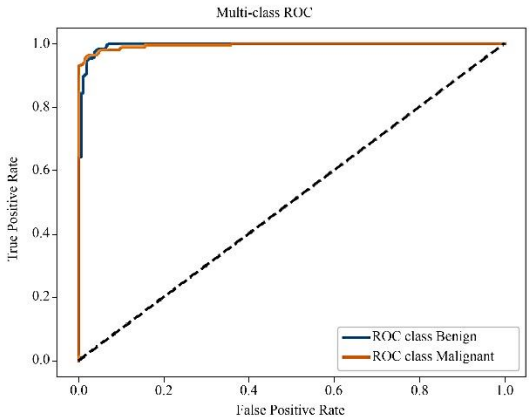

Fig. 5 ROC (proposed HCA)

Table 4. Overall comparison (accuracy, macro-F1, macro AUC, κ)

Model	Accuracy	Macro-F1	Macro AUC (OvR)	Cohen's kappa
Cross-Attn (no bias)	0.9485	0.9483	0.9951	0.8968
Self-Attn (stacked)	0.9508	0.9507	0.9918	0.9015
Proposed HCA (ours)	0.9687	0.9687	0.9956	0.9373

The new HCA (directional handcrafted CNN cross-attention with learnable bias prior B over (H×C) map) substantially improves balance and agreement measures (see Table 3 & Figure 4: Confusion Matrix Proposed HCA). HCA minimises the Benign Malignant flips from 20 (Cross-Attn) to 5, leading to Malignant precision increase from 0.919 to 0.978 while maintaining Malignant recall to 0.961. On the other hand, Benign recall increases to 0.977 with a Benign precision of 0.959 and specificity of about 0.961.

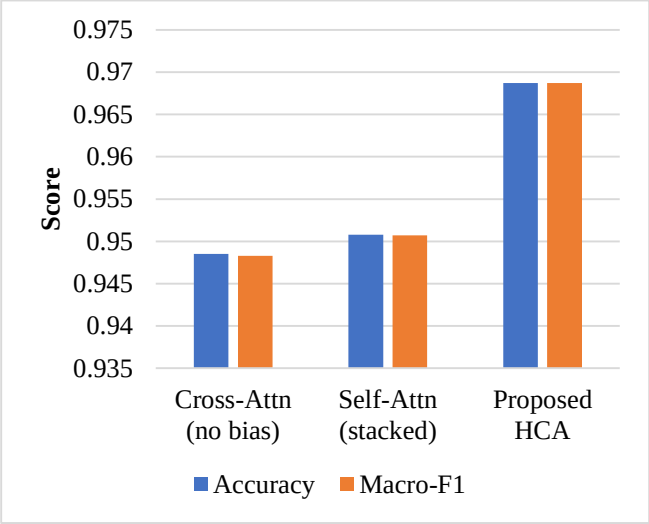

Fig. 6 Accuracy & macro-F1 by model

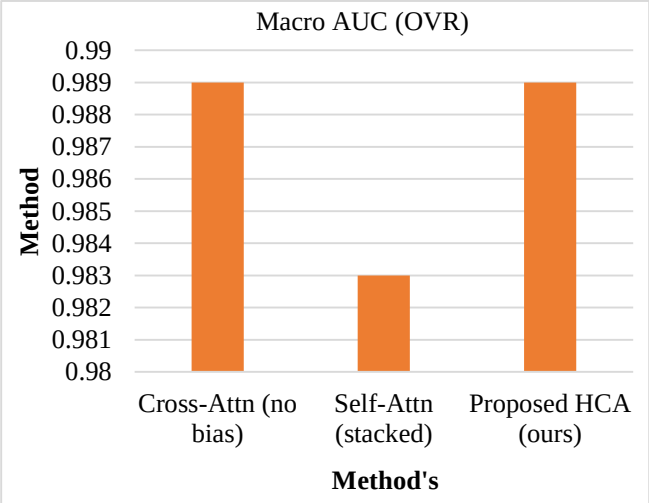

Fig. 7 Macro AUC by model

These coupled gains translate to the best overall outcomes: Accuracy = 0.9687, Macro-F1 = 0.9687, κ = 0.9373, and AUCmacro = 0.9956 (Table 4). The ROC characteristics in Figure 5: Multi-class ROC—Proposed HCA confirm near-perfect separability for both classes; the small yet consistent AUC edge over Cross-Attn (0.9956 vs 0.9951) is amplified in practice because the bias-guided attention aligns the score distribution with the operating threshold, converting separability into fewer cross-class errors, notably, far fewer false Malignant alarms.

Comparing these more directly, Figure 6 plots Accuracy and Macro-F1 across models, showing that the proposed method leads the two baselines by ~1.8–2.0 percentage points absolute for Cross-Attn and ~1.7–1.8 percentage points for Self-Attn. Figure 7 compares AUCmacro, where all models are good, but HCA still has the best value, suggesting that its attention prior does not come at the cost of poor separability for improved calibration.

Notably for clinical reproducibility, Figure 8 reveals Cohen's κ, where HCA has a scoring value of 0.937, which is markedly better than 0.897 (Cross-Attn) and 0.902 (Self-Attn). These are collectively addressed by the confusion level story: Cross-Attn's lack of prior causation for over-attending regions that resemble malignant CTS, over-calling Malignant on challenging Benign samples; Self-Attn's uncued combination lessens this propensity but lacks handcrafted-to-spatial supervisory signals; HCA reverses the need for directionality and adds trainable prior B, normalising which handcrafted channels pay attention to which spatial tokens. This results in an honed decision boundary that preserves Malignant expertise while not dampening imposter Malignant calls for Benign sections.

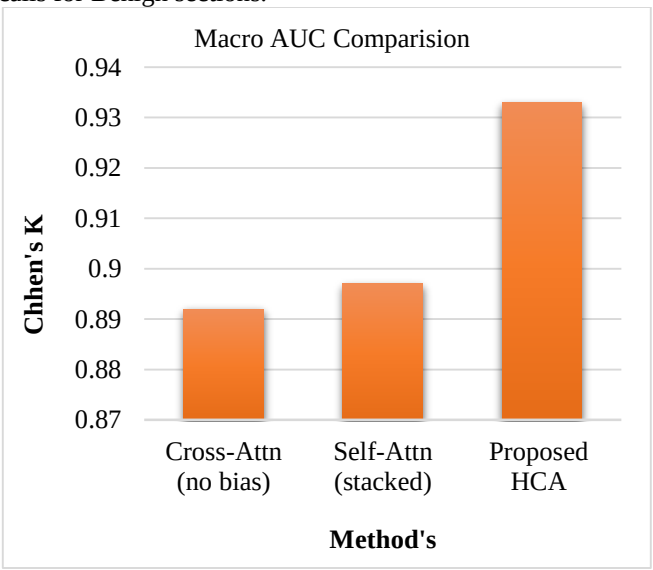

Fig. 8 Cohen's κ by model

4. Future Scope

Although the proposed Handcrafted-Guided Cross-Attention (HCA) framework is shown to provide outstanding and steady improvements over the baseline attention mechanisms in the context of the BreaKHis data set, some avenues remain unexplored that will be the focus of future work. First, while patient-level data sets split over different magnifications were used, validation over whole-slide image data sets coming from multiple institutions will help to confirm the robustness of different stain approaches and scanning devices. Second, while the proposed handcrafted guidance is based on small RGB histogram tokens, more informative pathology-oriented handcrafted priors, such as nuclei density statistics, gland shape characteristics, or hematoxylin-eosin components obtained after stain deconvolution, may be used in the future to improve further the proposed framework, which is advantageous in terms of interpretation and robustness. Third, the proposed bias-aware cross-attention module may also be embedded into pipelines that are based on hierarchical multiple-instance learning to enable the scalable application of the proposed approach in the context of gigapixel whole-slide images, while keeping the benefits of token-level merging. Finally, a further extension towards a task framework, where subtype classification is performed in parallel with the estimation of the prognosis risk, will represent an outstanding goal in the development of suitable computational pathology packages.

5. Conclusion

This paper proposes a bias-aware, directional, handcrafted-to-CNN cross-attention mechanism that fuses compact, interpretable colour-histogram tokens with spatial CNN tokens before pooling. By assigning fixed roles to the two family features-hand-crafted tokens as queries and CNN tokens as keys and values-and by learning an attention bias that represents priors linking stain and color statistics to morphologic regions, the approach renders representational separability into fewer clinically impactful errors. Relative to effective controls using cross-attention alone without preceding and cumulative self-attention, the design presented achieved more symmetrical class behaviour, more reliable decisions, and a more pronounced above-chance agreement with enhanced transparency while remaining lightweight, modular, and easy to integrate with typical backbones. Explicit use of an interpretable, handcrafted, small token set adds to further transparency, enabling attention-level inspection of spatial focus driven by color cues. Empirically, this recipe for histopathology from the study puts together a declarative fusion at token levels before any global summarization, complementary streams through dual pooling, injection of a trainable before early stabilization, and biasing of attention toward diagnostically stable tissue patterns. Each of these components contributes toward robust performance in realistic stain and magnification variability and dampens overcalling malignancy on challenging benign slides.

Acknowledgements

The authors thank the individuals behind the BreakHis dataset for their efforts and the institutions that host this data. Datasets of histopathology images, such as BreakHis, allow apples-to-apples comparisons among methods to facilitate reproducible computational pathology research.

Dataset Usage Statement

The BreakHis histopathology image dataset was exclusively used for research purposes from a publicly available source. It could be accessed from either the UFPR Visual & Recognition Lab site or from its mirror at Kaggle. It may be noted that according to the official site of UFPR Visual & Recognition Lab, BreakHis datasets are licensed under the Creative Commons Attribution 4.0 International license. The requirements of attribution, as stated on BreakHis and the platform's usage rules, were duly followed for utilising. The dataset was not used in clinical practice or the collection of protected health information. Such reuse or redistribution would be governed by BreakHis CC BY 4.0 license recommendations or Kaggle's Terms of Use as applicable. [11, 12].

Author Contribution

Subramanyam P. and Hari Krishnan G. conceptualised the study design and research plan as a whole. Subramanyam P. was involved in the design and implementation of the Handcrafted-aware Cross-Attention architecture, the setting up of the study design, and the creation of the handcrafted tokens. Additionally, he conducted the statistical analysis of the study design as well as the creation of the tables and figures. Hari Krishnan G. enhanced the study design and research methodology. He was also involved in the supervision of the study design results. Both authors were involved in the study design and results analysis. Additionally, they signed the final form of the manuscript.

References

[1] Fabio A. Spanhol et al., "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455-1462, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Soham Chattopadhyay et al., "DRDA-Net: Dense Residual Dual-Shuffle Attention Network for Breast Cancer Classification using Histopathological Images," *Computers in Biology and Medicine*, vol. 145, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Cong Cong et al., "Colour Adaptive Generative Networks for Stain Normalisation of Histopathology Images," *Medical Image Analysis*, vol. 82, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Musa Adamu Wakili et al., "Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-31, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Tania Afroz Toma et al., "Breast Cancer Detection Based on Simplified Deep Learning Technique with Histopathological Image Using BreaKHis Database," *Radio Science*, vol. 58, no. 11, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[6] Giulia Lucrezia Baroni, "Optimizing Vision Transformers for Histopathology: Pretraining and Normalization in Breast Cancer Classification," *Journal of Imaging*, vol. 10, no. 5, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[7] Abeer Heikal et al., "Fine Tuning Deep Learning Models for Breast Tumor Classification," *Scientific Reports*, vol. 14, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[8] Niful Islam et al., "Fusing Global Context with Multiscale Context for Enhanced Breast Cancer Classification," *Scientific Reports*, vol. 14, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[9] Lama A. Aldakhil, Haifa F. Alhasson, and Shuaa S. Alharbi, "Attention-Based Deep Learning Approach for Breast Cancer Histopathological Image Multi-Classification," *Diagnostics*, vol. 14, no. 13, pp. 1-27, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[10] Chengyang Gao et al., "Transformer based Multiple Instance Learning for WSI Breast Cancer Classification," *Biomedical Signal Processing and Control*, vol. 89, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[11] Xunping Wang, and Wei Yuan, "Nuclei-Level Prior Knowledge Constrained Multiple Instance Learning for Breast Histopathology Whole Slide Image Classification," *iScience*, vol. 27, no. 6, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[12] Lu Cao et al., "Multi-Branch Spectral Channel Attention Network for Breast Cancer Histopathology Image Classification," *Electronics*, vol. 13, no. 2, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[13] Rui Ding et al., "A Deep Multi-Branch Attention Model for Histopathological Breast Cancer Image Classification," *Complex & Intelligent Systems*, vol. 10, pp. 4571-4587, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[14] Aadhi Aadhavan Balasubramanian et al., "Ensemble Deep Learning-Based Image Classification for Breast Cancer Subtype and Invasiveness Diagnosis from Whole Slide Image Histopathology," *Cancers*, vol. 16, no. 12, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[15] Xiaoli Wang et al., "MFF-ClassificationNet: CNN-Transformer Hybrid with Multi-Feature Fusion for Breast Cancer Histopathology Classification," *Biosensors*, vol. 15, no. 11, pp. 1-29, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[16] Venkata Nagaraju Thatha et al., "Histopathological Image based Breast Cancer Diagnosis using Deep Learning and Bio Inspired Optimization," *Scientific Reports*, vol. 15, pp. 1-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[17] Guolan Wang et al., "Multi-Classification of Breast Cancer Pathology Images based on a Two-Stage Hybrid Network," *Journal of Cancer Research and Clinical Oncology*, vol. 150, pp. 1-15, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[18] Oluwatosin Tanimola et al., "Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images," *Analytics*, vol. 3, no. 4, pp. 461-475, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[19] Suxing Liu, and Byungwon Min, "DCS-ST for Classification of Breast Cancer Histopathology Images with Limited Annotations," *Applied Sciences*, vol. 15, no. 15, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[20] Faseela Abdullakutty et al., "Histopathology in Focus: A Review on Explainable Multi-Modal Approaches for Breast Cancer Diagnosis," *Frontiers in Medicine*, vol. 11, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[21] Yuting Yan et al., "Breast Cancer Histopathology Image Recognition using Transformer Learning," *Medical Engineering & Physics*, vol. 138, no. 1, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[22] Dehua Liu et al., "Dual-Attention Multiple Instance Learning Framework for Pathology Whole-Slide Image Classification," *Electronics*, vol. 13, no. 22, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[23] Martim Afonso et al., "Multiple Instance Learning for WSI: A Comparative Analysis of Attention-Based Approaches," *Journal of Pathology Informatics*, vol. 15, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[24] Attasuntorn Traisuwan et al., "Color Normalization in Breast Cancer Immunohistochemistry Images Based on Sparse Stain Separation and Self-Sparse Fuzzy Clustering," *Diagnostics*, vol. 15, no. 18, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[25] Qinyi Zhang et al., "Enhanced Nuclear Information Fusion and Visual Transformer for Pathological Breast Cancer Image Classification," *Scientific Reports*, vol. 15, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[26] Bin Yang et al., "Transformer-based Multiple Instance Learning Network with 2D Positional Encoding for Histopathology Image Classification," *Complex & Intelligent Systems*, vol. 11, pp. 1-17, 2025. [CrossRef] [Google Scholar] [Publisher Link]
[27] Jin-Gang Yu et al., "Prototypical Multiple Instance Learning for Predicting Lymph Node Metastasis of Breast Cancer from Whole-Slide Pathological Images," *Medical Image Analysis*, vol. 85, 2023. [CrossRef] [Google Scholar] [Publisher Link]