

SAWL-Net: A Statistical Attentions and Wavelet Aided Lightweight Network for Classification of Cancers in Histopathological Images

Surya Majumder , Aishik Paul , Friedhelm Schwenker , and Ram Sarkar 

Abstract—Addressing the formidable challenges posed by the diagnosis and management of various types of cancer, including breast, colon, lung, and colorectal cancer, demands innovative solutions to streamline histopathological analysis processes. In this study, we propose a novel lightweight convolutional neural network (CNN) called statistical attentions and wavelet aided lightweight network (SAWL-Net) architecture based on MobileNetV2 equipped with hybrid statistical similarity and wave format-aided attention mechanisms, specifically tailored to the demands of cancer histopathology. By leveraging the capabilities, our model incorporates a lightweight design while ensuring high-performance outcomes. We introduce a unique combination of Pearson correlation coefficient, Spearman rank correlation, and cosine similarity metrics, alongside a specialized wave conversion technique to enhance the detection of similarities across different channels of histopathological data, while providing a holistic approach to the model. In this study, we have considered breast, colorectal, and lung & colon cancer datasets for experimentation. Notably, our model surpasses prevailing state-of-the-art methodologies, showcasing its efficacy in optimizing diagnostic accuracy and expediting treatment strategies for varied cancer types. Our codes are publicly available at the GitHub repository.

Impact Statement—Our research enhances cancer classification accuracy in a MobileNetV2 based neural network. The Pearson correlation coefficient layer introduces attention weights in the outermost skip connection, prioritizing relevant low-level features. The Spearman rank correlation layer introduces attention weights in the innermost skip connection, prioritizing among high-level features. The wave conversion layer allows for localised pattern recognition in a more sequential approach across all channels simultaneously, giving a holistic view, which in turn produces more accurate classification results. “Low-level

features” refer to early convolutional layers capturing simple patterns (edges, corners), while linear and nonlinear relationships pertain to feature map dependencies.

Index Terms—Breast cancer, colon cancer, colorectal cancer, cosine similarity, histopathological image, lung cancer, Pearson correlation coefficient, spearman rank correlation, similarity attention.

I. INTRODUCTION

HISTOPATHOLOGICAL images serve as indispensable tools in clinical diagnostics, particularly in the realm of cancer assessment, where a meticulous examination of cellular structures and characteristics is pivotal for subtype identification, staging, and genetic profiling. The precise classification of cancerous categories within stained pathological images stands as a cornerstone in pathological research, yet poses significant time constraints for pathologists, particularly given the dense nature of whole-slide images. Automating this classification process is imperative, especially considering the variability in the process due to visibility and large details, which remains a persistent challenge in clinical pathology analysis. As highlighted in a study by Metter et al. [1], the current availability of pathologists often struggles to meet the escalating demand for accurate and timely pathological evaluations.

Recent advancements in machine learning and deep learning have revolutionized histopathological image classification. Wang et al. [2] proposed convolutional and recurrent neural networks with short-term and long-term spatial correlations for efficient breast cancer histopathological image classification. Masaud et al. [3] proposed ensemble-based multitissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework using dilated ResNet and deep support vector machine (SVM). Transformer based colonoscopy image classification and retrieval has also been used for detecting microsatellite instability in colorectal cancer [4]. Order-ViT [5] proposed a vision transformer for order learning in cancer classification in pathology images. Supervised contrastive learning has been used for the classification of colon cancer [6]. MDFF-Net [7] proposed a multiple dimensional feature fusion network for breast histopathology image classification. Guangli et al. [8] proposed a lung, colon, and breast histopathological image classification technique using embedded fusion mutual learning. Taheri et al. [9]

Received 5 October 2024; revised 28 November 2024; accepted 25 January 2025. Date of publication 30 January 2025; date of current version 31 July 2025. This article was recommended for publication by Associate Editor Feng Xia upon evaluation of the reviewers' comments. (Corresponding author: Friedhelm Schwenker.)

Surya Majumder is with the Computer Science and Engineering, Heritage Institute of Technology, Kolkata 700107, India (e-mail: suryamajumder0802@gmail.com).

Aishik Paul, Friedhelm Schwenker, and Ram Sarkar are with the CMATER Research Lab, Jadavpur University, Jadavpur 700032, India (e-mail: aishik.paul.ece26@heritageit.edu.in; friedhelm.schwenker@uni-ulm.de; rsarkar@ieee.org).

Digital Object Identifier 10.1109/TAI.2025.3536424

proposed four methods for breast cancer diagnosis across binary versus multiclass and magnification-specific (MS) versus magnification-independent (MI) classification scenarios. They developed the multilevel feature fusion (MLF2) model, an automatic score-level fused convolutional neural network (CNN) based on a pretrained DenseNet-121. MLF2-CNN integrates feature extraction and classification, fusing multiple classifiers trained on multi-level features to produce final predictions. Additional connections in the CNN architecture capture multi-stage features for more comprehensive diagnosis.

Addo et al. [10] proposed BCHI-CovNet, a lightweight CNN model for classifying histopathological breast images. They introduced a multiscale depth-wise separable convolution to capture both low- and high-resolution patterns and a pooling module with multihead self-attention to capture second-order statistics and long-range pixel dependencies. This design reduced computational complexity, making it suitable for resource-limited medical devices. BCHI-CovNet achieved a high accuracy on the BreakHis dataset (up to 99.22%) and the BACH dataset (99.38%). In another lightweight approach, Varun et al. [11] proposed a CAD system for early breast cancer detection using a stacked ensemble-based deep CNN model. The ensemble includes MobileNetV2, ShuffleNet, and SqueezeNet, trained on the BreakHis dataset to classify benign and malignant tissues from histopathological images. The final predictions were made through an average voting-based classifier. The stacked model outperformed individual CNNs, improving accuracy and robustness in breast cancer diagnosis. Similarly, Saha et al. [12] proposed Breast-NET, a deep CNN model for breast cancer detection and grading using histological images. Evaluated on the BreakHis and IDC datasets, the model showed a good generalization and efficiency, validated through extensive experiments and ablation studies. Additionally, transfer learning with seven pre-trained CNNs enhanced the performance. Breast-NET outperformed existing methods in accuracy, space, and computational complexity across multiple datasets.

Attention mechanisms are a technique to spotlight particular image regions, enhancing model performance and resilience. Dimension information embedding attention network (DIEANet) [13] has been proposed for lung adenocarcinoma histopathological image grading. Provath et al. [14] used a global context attention based CNN model for classification of lung and colon cancer. A multilevel convolutional and attention learning was proposed by Dabass et al. [15] for classification of cancer grades and tissue structures in colon cancer. The method, called IL-MCAM [16], used a multiple channel attention and interactive learning based mechanism for colorectal cancer classification. TransSurv [17] used a cross-modal fusion transformer survival analysis model, based on cross attention, using histopathological images and genomic data for classification of colorectal cancer. Yan et al. [18] proposed a guided attention based model for breast cancer classification. A spatial attention system for pathological complete response prediction for breast cancer with serial histopathology images in multiple stains was proposed by Duanmu et al. [19]. While no mechanisms are standalone enough to achieve great results on

complex histopathological images across various cancer types, the approach of attention mechanism becomes prominent due to its ability to highlight features that improve the model's performance.

To optimize integration and maximize benefits from diverse methods, a channel-wise hybrid similarity attention mechanism is essential. Operating across multiple levels, this mechanism must adeptly grasp semantic nuances while filtering out irrelevant features to enhance model clarity. With these objectives in mind, the main contributions of this work can be summarized as follows.

- 1) Proposed a lightweight neural network model for classification of histopathological images, which utilizes the MobileNetV2 model as a backbone. Size of model is pretty low even after adding all accessories.
- 2) A Pearson correlation coefficient with cosine similarity based attention mechanism is introduced to highlight similarities among low-level features in the model, while capturing both linear and nonlinear relationships.
- 3) A Spearman rank correlation based attention mechanism, fused with cosine similarity, is introduced to highlight similarities among high-level features, besides providing a holistic view of the similarity among feature vectors, leveraging both rank-based and direction-based measures.
- 4) Proposed a unique way of treating channel outputs of a layer as wavelets or continuous sequences of data for finding local patterns in them.
- 5) Achieved superior results on standard datasets, namely BreakHis, NCT-CRC-HE-100K, and LC25000.

While attention mechanisms have been commonly used, our contribution lies in the innovative use of Pearson and Spearman correlations in tandem with cosine similarity and wavelet features, capturing both linear and nonlinear relationships.

II. DATASET DESCRIPTION

We have considered three different histopathological images datasets from different domains such as breast cancer, colorectal cancer, and lung & colon cancer, as it ensures robustness and generalization of the proposed model by exposing it to a wide range of pathological variations, aiding in comprehensive feature learning and improved performance across various cancer types.

- 1) *BreakHis*: A vast collection of histopathological images of breast tumor tissues is available for breast cancer research [20], encompassing both benign and malignant tumors categorized into four main classes, *Benign*: adenosis (AD), fibroadenoma (FA), phyllodes tumor (PT), tubular adenoma (TA); *Malignant*: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). The images are available at various magnifications, including 40x, 100x, 200x, and 400x. Fig. 1 displays the internal relations.
- 2) *NCT-CRC-HE-100K*: A collection of 100 000 distinct image patches from Hematoxylin & Eosin (H&E) stained histological images of human colorectal cancer (CRC)

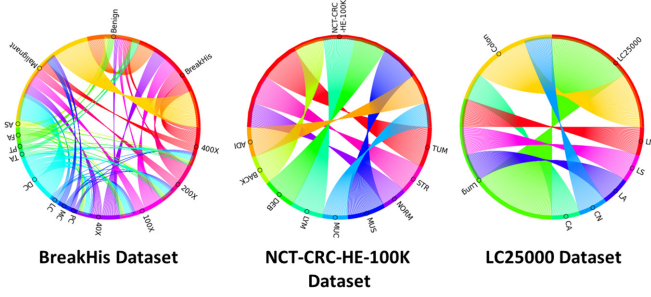


Fig. 1. Chord diagram of each dataset highlighting their internal relationships.

and normal tissue [21]. The tissue classes are adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM). Fig. 1 displays the internal relations.

- 3) *LC25000*: LC25000 lung and colon histopathological image dataset [22], contains color 25 000 images with five classes of 5000 images each. It has two main classes with subclasses: colon [colon_aca (CA) for adenocarcinomas, colon_n (CN) for benign tissues] and lung [lung_aca (LA) for adenocarcinomas, lung_scc (LS) for squamous cell carcinomas, lung_n (LN) for benign tissues]. Fig. 1 displays the internal relations.

A. Data Preprocessing

In preprocessing, all images of all three datasets, have been resized to shape 256×256 . These images were then normalized using standard normalization i.e., dividing by 255. The datasets were then split into train-validation-test sets in a 7:1:2 ratio for all subsequent analysis, enhancing accuracy and interpretability of medical imaging data.

III. PROPOSED METHODOLOGY

Our proposed model, called statistical attentions and Wavelet aided lightweight network (SAWL-Net), is tailored to efficiently manage histopathological datasets of diverse complexities with increased adaptability, all while maintaining an exceptionally small size of just 8.87 MB. Leveraging MobileNetV2 [23] as the backbone, an image classification model renowned for its proficiency in capturing multi-scale contextual details, further enhances its capabilities. The architecture enhances efficiency with linear bottlenecks, inverted residuals, and depth-wise separable convolution. It includes two types of blocks—shrinking blocks with a stride of two and residual blocks with a stride of one, each with three levels. Utilizing a width multiplier, it adjusts for hardware limitations. We have used the default width multiplier, $\alpha = 1.0$ in our work here. To enhance the performance of the backbone, we have implemented a similarity score-based attention mechanism that considers uncertainty in feature maps' predictions and assigns different importance levels using the Pearson correlation coefficient (PCC) at one stage

and Spearman rank correlation (SRC) at another, to highlight the important spatial regions for accurate classification. This, in turn, allows the model to focus on crucial features. A block diagram of the proposed model is presented in Fig. 2.

A. Implementation of the Entire Model

Our model is created using the MobileNetV2 model, as our base model. We have employed the PCC and SRC function-based attention mechanisms that take the entire output of their respective previous layer, dimension $[\text{batch} \times \text{height} \times \text{width} \times \text{channels}]$. Using the statistical methods and a re-parameterized exponential function, we generate scores for every channel and use it in our attention mechanism. However, the processing of the inputs and calculation of the final attention scores are different in different scenarios. The original MobileNetV2 architecture only contains residual connections, but here we inserted two skip connection on which PCC and SRC lie.

Significance of Pearson Correlation Coefficient for Score Calculation: Our model leverages a novel layer for feature score calculation based on PCC within the outer skip connection to provide importance to feature maps. This layer aggregates input feature maps using a pixel-wise spatial averaging and computes PCC scores and cosine similarity (CS) scores between each input feature map and the aggregated one

$$\bar{\mu} = \frac{1}{C} * \sum_{i=1}^C FM_i. \quad (1)$$

The aggregated feature map ($\bar{\mu}$) is generated by averaging each feature map pixel-wise, as described in (1), where C = No. of channels (feature maps), FM = input tensor, and FM_i represents a feature map of the i th channel of FM . Then, PCC and CS scores are computed between each extracted feature map and the aggregated feature map using (2) and (3), respectively, where $\bar{\mu}$ = aggregated feature map, \vec{x} = the feature map for score calculation, $\mu_{\vec{f}}$ and $x_{\vec{f}}$ are their equivalent flattened versions respectively, n = no. of elements in the feature map, μ_{μ} and x_{μ} are the mean of the elements of the feature map μ and x , respectively

$$\text{PCC}(\bar{\mu}, \vec{x}) = \frac{\sum_{i=1}^n (\mu_i - \mu_{\mu}) * (x_i - x_{\mu})}{\sqrt{\sum_{i=1}^n (\mu_i - \mu_{\mu})^2 * (x_i - x_{\mu})^2}} \quad (2)$$

$$\text{CS}(\bar{\mu}_{\vec{f}}, \vec{x}_{\vec{f}}) = \frac{\sum_{i=1}^n \mu_{\vec{f}_i} * x_{\vec{f}_i}}{\|\mu_{\vec{f}}\| * \|x_{\vec{f}}\|}. \quad (3)$$

The scores calculated using (2) and (3) are then weighted averaged to get the final similarity scores (FSS₁)

$$\begin{aligned} \text{Scores} &= e^{a * (\text{FSS}_1 - h)} * c + k \\ a &= 3.1; h = 1.065; c = 1.1; k = 0.1. \end{aligned} \quad (4)$$

Using the reparameterized exponential function depicted in Fig. 3 and (4), we interpolate the resultant score between 0 and 1 for multiplication with channels as attention, the values of a , h , c , and k were determined so that no feature maps attention becomes zero or tends to zero and the feature maps with higher

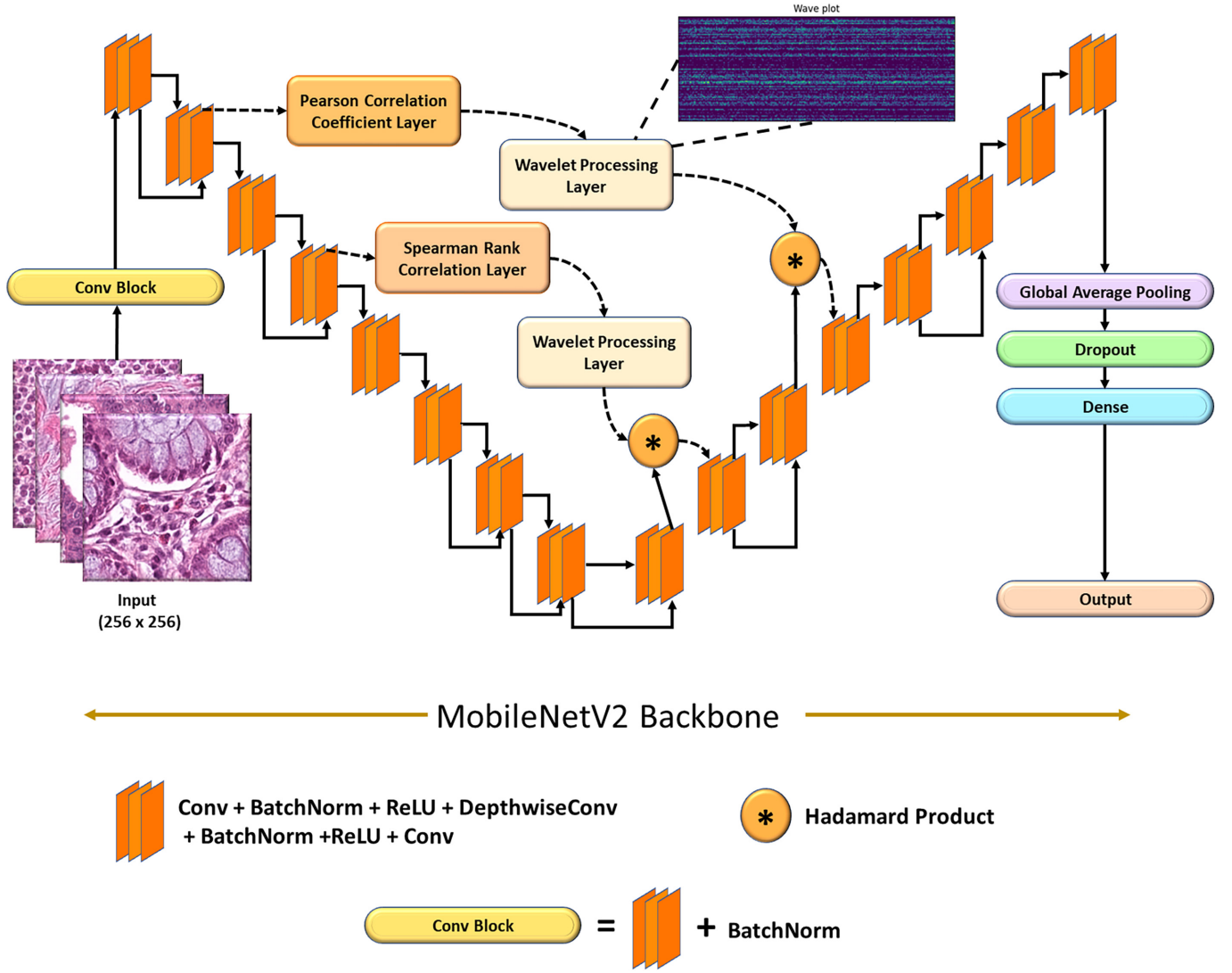


Fig. 2. Architecture of the proposed SAWL-Net model.

similarity gets more attention. The parameter values in (4), are determined mathematically as well as experimentally. It is determined from the concept of shearing in linear algebra, where a function $f(x)$ has to be sheared so that it passes through two specific points here, $(0, 0.141)$ and $(1, 0.999)$. That is because the model layer output cannot be greater than 1 and less than 0, as the original input has been normalized, and there are multiple batchnormalization layers inside the MobileNetV2 architecture; while the rank also has to remain between the already specified range. We first reparameterized e^x to pass through $(0, 0.5046)$ and $(0.8, 1)$. Therefore, we reparameterized function using variables c and k . Let $f(x)$ be our function e^x . Now using shearing we get our desired function by $g(x) = A * f(x) + B$. Solving for the two specified points and on obtaining the values of A and B , we would get $A = 1.1$ and $B = 0.1$ making $g(x)$ the desired reparameterized function. These A and B , become our c and k variables. Later on keeping normalization in mind, we added the a and h parameters to adjust the graph by moving the curve vertically and horizontally so that the minimum attention given

is 0.141 or 14% and max attention as 0.999 tending to 1, which is to bring function to pass through $(0, 0.141)$ and $(1, 0.999)$. This results in $a = 3.1$ and $h = 1.065$. As in the outer skip connection, the model is likely to have captured higher-level features that represent complex relationships between various histopathological structures or patterns, they may exhibit linear dependencies or correlations. For instance, certain types of cell arrangements or tissue structures may exhibit linear relationships in their distribution or arrangement. Mixing PCC with CS provides a more comprehensive understanding of similarity, incorporating both linear and nonlinear aspects. Algorithm 1 provides a detailed approach.

Significance of Spearman Rank Correlation for Score Calculation: Another novel layer our model leverages for feature score calculation is based on SRC (ρ) within the inner skip connection to provide importance to feature maps (this layer is referred to as SRCC later on). As the inner skip connection likely operates at a lower level, capturing more granular details or local patterns within the images, the relationships between

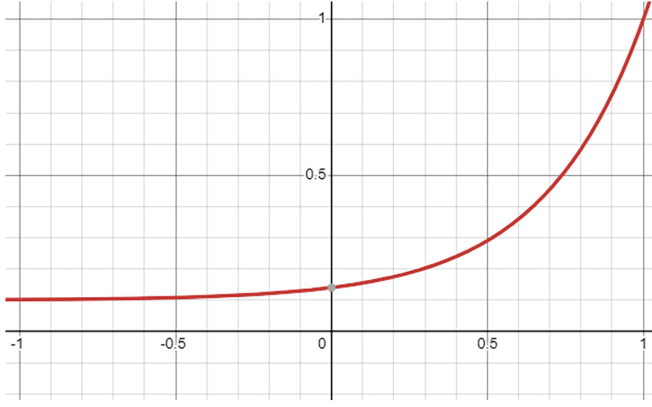


Fig. 3. Modified exponential function used in the present work.

features may be more complex and nonlinear. For example, the arrangement of individual cells or small tissue structures may not follow a linear pattern but could still have meaningful relationships. ρ captures the monotonic relationships between them without assuming linearity. Using (1), we again calculate the aggregated feature map μ , after which using (5) and (3), we calculate the ρ and CS scores, respectively, which are then weighted averaged before being interpolated using (4) to get scores FSS_2 that are multiplied to the input as attentions. Here also $\vec{\mu}$ = aggregated feature map, \vec{x} = the feature map for score calculation, and n = no. of pixels in the feature map

$$\rho(\vec{\mu}, \vec{x}) = 1 - \frac{6 * \sum_{i=1}^n (\mu_i - x_i)^2}{n * (n^2 - 1)}. \quad (5)$$

Using SRC alongside CS ensures that both the local, nonlinear relationships captured by Spearman's correlation and the directional similarities captured by CC are accounted for. Algorithm 1 depicts the entire process in details. Fig. 4 shows the effect of the attentions on feature maps.

Significance of Wave Format for Localized Pattern Identification: Converting the correlation outputs to wave format allows for capturing localized patterns within the data. Waves inherently represent variations in the data over time or space, enabling Conv1D to detect patterns that occur at specific locations within the input. By processing the flattened and concatenated data as a wave, we enable the network to capture patterns that might exist across spatial dimensions and channels in a more efficient manner. The original input to the PCC layer and the SRCC layer are then respectively multiplied with the output of their corresponding wave format results of Conv1D blocks, which has been reshaped to Hadamard product with corresponding layer output to join skip connection with; elementwise so that proper pixels now get proper attention based on the previous operations. Fig. 5, shows some sample images of different classes and their corresponding amplitude wave maps. Algorithm 1 depicts the same in detail.

IV. RESULTS AND DISCUSSION

For all datasets, we have used the original image size of 256×256 as the input and trained the model for 200 epochs.

Algorithm 1: Algorithm of the SAWL-Net Model.

```

 $N \leftarrow$  Total no. of feature maps
 $mode \leftarrow$  String
 $\alpha \leftarrow$  value
function COSINE_SIMILARITY( $x\_flat, y\_flat$ )
     $score \leftarrow$  Equation 3  $\triangleright$  Calculating CS score between
    flattened x and y.
    return score
end function
function PCC( $x\_flat, y\_flat$ )
     $CS \leftarrow$  cosine_similarity( $x\_flat, y\_flat$ )
     $PCC \leftarrow$  Equation 2  $\triangleright$  Calculating PCC score
    between x and y.
     $score \leftarrow \frac{(\alpha * PCC) + CS}{2.0}$ 
    return score
end function
function SRCC( $x\_flat, y\_flat$ )
     $CS \leftarrow$  cosine_similarity( $x\_flat, y\_flat$ )
     $\rho \leftarrow$  Equation 5  $\triangleright$  Calculating  $\rho$  score between x and
    y.
     $score \leftarrow \frac{(\alpha * \rho) + CS}{2.0}$ 
    return score
end function
function SIMILARITY_COMPUTE( $mean, argv$ )
     $similarities \leftarrow$  array()
    for  $arg_i \in argv$  do
        if  $mode == "pearson"$  then
             $sim \leftarrow$  PCC(Flatten(mean), Flatten( $arg\_i$ ))
        else if  $mode == "spearman"$  then
             $sim \leftarrow$  SRCC(Flatten(mean), Flatten( $arg\_i$ ))
        end if
         $similarities \leftarrow$  Add sim to this array
    end for
     $fscore \leftarrow$  Equation 4
     $res = argv * fscores$   $\triangleright$  Applying attention
    return res
end function
function CALL( $inputs$ )
     $L \leftarrow$  array( $inputs[:, :, i]$ )  $\triangleright$  Extracting feature maps
    from the input
     $agg\_feature\_map \leftarrow$  Equation 1
     $results \leftarrow$  similarity_compute( $agg\_feature\_map, *L$ )
     $\triangleright$  Passing mean and unpacked list as a function input
     $res \leftarrow$  array()
    for  $i \leq results.shape[-1]$  do
         $res \leftarrow$  Add Flatten( $results[:, :, i]$ )
         $\triangleright$  Converting channels to wave format
    end for
    return res  $\triangleright$  res is returned and passed through
    Conv1D layers
end function
 $results \leftarrow$  Call( $inputs$ )

```

We have employed a 70-20-10% train-validation-test split, a learning rate of 0.00001, the Adam optimizer, and a batch size of 16. During training, we have utilized the Sparse categorical

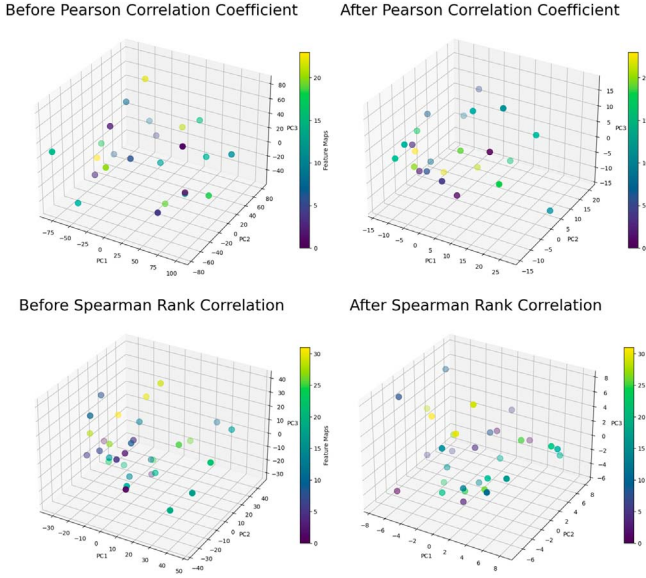


Fig. 4. PCA based representation of the effect of attention on feature maps; PCA applied on feature maps before plotting each feature map.

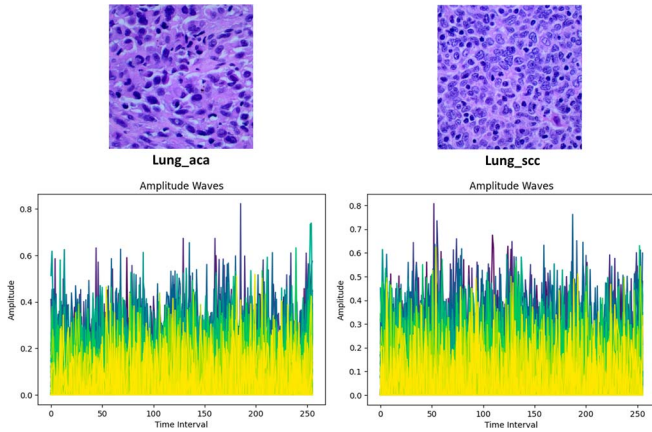


Fig. 5. Sample images of two lung classes and their corresponding amplitude wave maps; the data of each channel is treated as amplitude and plotted.

cross entropy loss. Model evaluation is conducted using accuracy (acc.), precision (pre.), recall (rec.), and F1 score. Figs. 6 and 7 depict the confusion matrix and ROC-AUC curves obtained by our model on each dataset, highlighting the performance of our model.

A. Efficiency Results and Choice of Backbone

The selection of the backbone architecture for SAWL-Net was guided by the need to achieve an optimal balance between model efficiency and performance across diverse datasets. To ensure lightweight design and scalability, MobileNetV2 was adopted as the backbone, reducing the number of parameters in our SAWL-Net model to 2.32M and model size to 8.87 MB as shown in Table I. This is notably smaller than traditional backbones such as Xception, DenseNet201, and ResNet50, which have parameter counts exceeding 20M and model sizes well

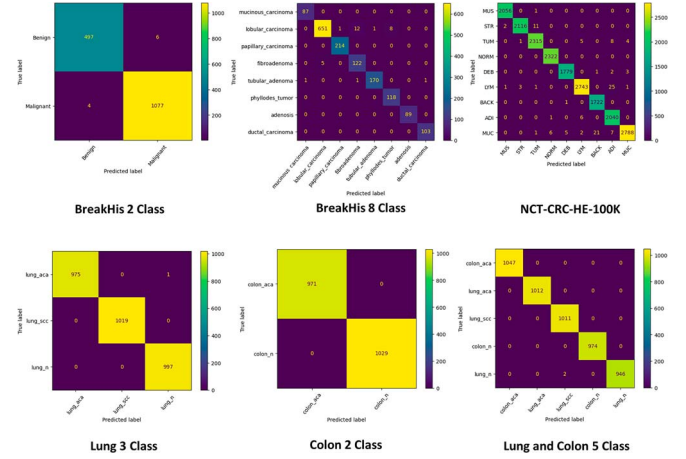


Fig. 6. Confusion matrices obtained by the SAWL-Net on each dataset.

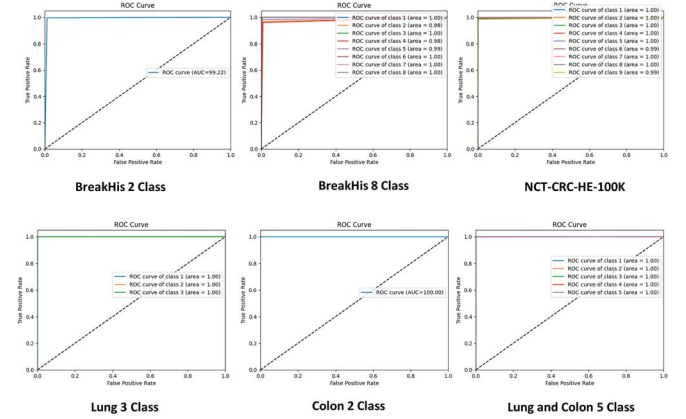


Fig. 7. ROC-AUC curves obtained by the SAWL-Net on each dataset.

over 80 MB. Additionally, the inference time on GPU, as shown in Table I for SAWL-Net (3.99 ms) is on par with MobileNetV2 (3.8 ms) and substantially faster than other backbones.

In terms of classification performance, SAWL-Net consistently outperforms all competing models across all benchmark datasets considered here. It achieves state-of-the-art (SOTA) accuracies on all the datasets, which have been depicted in details later. Specially, on BreakHis-2-Class (99.37%) and BreakHis-8-Class (98.11%) datasets, it significantly surpasses other backbones. Notably, for the NCT-CRC-HE-100K dataset, SAWL-Net achieves 99.40% accuracy, outperforming the closest competitor, DenseNet201, which achieved 95.96%. Similarly, on the LC25000 datasets (Lung, Colon, and combined), SAWL-Net delivers near-perfect results, with 99.97% accuracy on LC25000-Lung and 100.0% on LC25000-Colon datasets.

A comprehensive ablation study provided later reveals that SAWL-Net consistently leads across most evaluation metrics, including precision, recall, and F1 score, indicating its robustness and reliability. This demonstrates the effectiveness of MobileNetV2 as a lightweight backbone and highlights SAWL-Net's ability to leverage its architecture to achieve superior performance while maintaining computational efficiency.

TABLE I
COMPARISON OF EFFICIENCY AND CLASSIFICATION ACCURACY OF THE PROPOSED MODEL WITH SOME WELL-KNOWN CNN MODELS

Models	Xception	DenseNet201	ResNet50	InceptionV3	NASNetMobile	EfficientNetB0	MobileNetV2	SAWL-Net
No. of Parameters in Millions	22.9M	20.2M	25.6	23.9	5.4	5.3	2.27	2.32
Size in MB	88	80	98	92	23	29	8.66	8.87
Time (ms) per inference step (GPU)	8.1	6.7	4.6	6.9	6.7	4.9	3.8	3.99
Accuracy (%)	BreakHis-2-Class	97.69	95.91	90.73	96.83	95.17	96.94	99.37
	BreakHis-8-Class	85.69	85.94	61.74	79.25	75.75	70.26	98.11
	NCT-CRC-HE-100K	95.06	95.96	89.17	92.76	91.59	90.08	99.40
	LC25000-Lung	98.84	99.47	80.29	99.10	99.87	96.87	99.97
	LC25000-Colon	100.0	100.0	90.80	99.95	99.80	91.77	100.0
	LC25000-Lung & Colon	98.68	98.70	79.22	99.26	99.01	94.83	99.96

TABLE II
PERFORMANCE OF THE CLASSIFICATION MODELS, ABLATION STUDY ON EACH OF THE DATASETS

Model	BreakHis-2-Class				BreakHis-8-Class			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
(i)	80.14	83.67	84.04	83.85	77.31	75.97	89.48	80.32
(ii)	86.77	86.21	86.96	86.58	82.84	79.99	88.47	80.30
(iii)	88.91	89.35	89.68	89.51	83.76	82.11	89.35	81.37
(iv)	99.37	99.32	99.22	99.27	98.11	97.51	98.88	98.17
	NCT-CRC-HE-100K				LC25000-Lung			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
(i)	95.67	95.26	96.62	95.94	98.76	98.89	98.84	98.86
(ii)	97.85	97.80	97.99	97.89	99.26	99.12	99.20	99.16
(iii)	97.90	96.83	97.98	97.40	99.53	99.47	99.46	99.46
(iv)	99.40	99.38	99.47	99.42	99.97	99.97	99.97	99.97
	LC25000-Colon				LC25000-Lung & Colon			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
(i)	99.42	99.37	99.46	99.41	98.51	98.47	98.49	98.48
(ii)	99.87	99.72	99.86	99.79	99.20	98.93	99.11	99.02
(iii)	99.91	99.89	99.88	99.88	99.59	99.60	99.57	99.58
(iv)	100.0	100.0	100.0	100.0	99.96	99.96	99.96	99.96

Note: Bold indicates the results obtained by our model.

TABLE III
PERFORMANCE OF THE CLASSIFICATION MODELS ON EACH DATASET WHEN THE PCC AND SRCC ARE SWAPPED WITH WAVELET PROCESSING LAYER POSITIONS

Datasets	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
BreakHis-2-Class	82.14	81.93	83.23	82.57
BreakHis-8-Class	79.82	79.11	78.98	79.04
NCT-CRC-HE-100K	95.19	95.73	95.60	95.66
LC25000-Lung	99.40	99.33	99.51	99.42
LC25000-Colon	99.90	99.92	99.89	99.90
LC25000-Lung & Colon	98.99	99.17	98.90	99.03

B. Ablation Study and Comparison With SOTA

We have analyzed the datasets to determine the best setup and parameters for our model through an extensive ablation study. The experiments are as follows: 1) MobileNetV2 (backbone); 2) MobileNetV2 along with PCC layer; 3) MobileNetV2 along with SRC layer; and 4) the proposed model. Table II highlights the significant impact of the PCC and SRC on the performance enhancement of MobileNetV2. To experimentally prove that the positional orders, i.e., PCC followed by Wavelet processing layer, and SRCC followed by Wavelet processing layer, we have swapped their positions and conducted detailed experiments across all the datasets as highlighted in Table III. This table shows the lowered accuracy, precision, recall and F1 values indicating that our current placement of PCC followed by Wavelet processing layer, and SRCC followed by Wavelet processing layer, is the correct order to use.

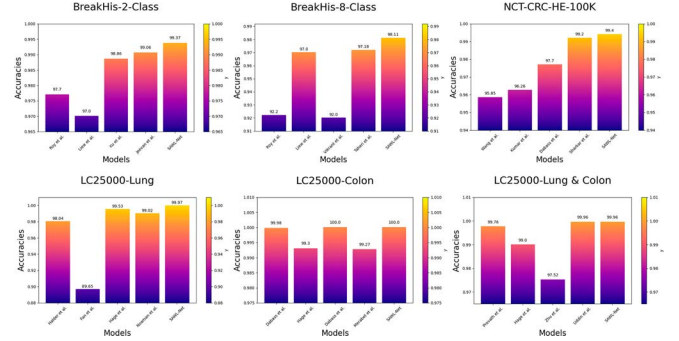


Fig. 8. Statistical view of the results of our model versus the SOTA results.



Fig. 9. Sample misclassified images by our SAWL-Net model.

We have conducted comparison between our approach and SOTA methods across all three datasets, and we have summarized the outcomes in Table IV. All results are in %. Bold figures highlight superior performance. The results illustrate that our model outperforms the SOTA methods across all three datasets. Fig. 8 provides a statistical view of the comparison of results among SOTAs.

All these SOTA comparisons highlight some great increase in performance results, such as in metrics such as accuracy, precision, recall, and F1 score, showcasing the fact that our model is highly efficient in classifying the different types of diseases.

Error Analysis: Fig. 9 depicts some images that our model fails to classify properly. One possible reason could be variations in staining protocols, leading to inconsistencies in image appearance that the model may not have been adequately trained to handle or the image quality might also be an issue. Either way it highlights the fact that our model is not 100% perfect and there is a scope for improvement in the future.

C. Visualization of Results

In this section, we take help from four data visualization tools, namely GradCAM, t-SNE plots, Activation Maps and

TABLE IV
PERFORMANCE COMPARISON OF THE PROPOSED MODEL WITH SOTA METHODS

Model	BreakHis-2-Class				Model	BreakHis-8-Class			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)		Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
Roy et al. [24]	97.70	97.00	98.40	97.70	Roy et al. [24]	92.20	89.80	85.40	87.60
Liew et al. [25]*	97.00	97.00	97.00	97.00	Liew et al. [25]*	97.00	97.00	96.87	96.88
Xu et al. [7]*	98.86	98.65	98.75	98.70	Vikranth et al. [26]	92.00	92.20	92.00	92.00
Jeevan et al. [27]	99.06	-	-	-	Taheri et al. [9]*	97.18	97.61	96.88	97.24
Proposed Method	99.37	99.32	99.22	99.27	Proposed Method	98.11	97.51	98.88	98.17
Model	NCT-CRC-HE-100K				Model	LC25000-Lung			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)		Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
Wang et al. [2]	95.85	-	-	95.82	Halder et al. [28]	98.04	99.12	98.33	98.72
Kumar et al. [29]	96.26	96.44	96.34	96.38	Fan et al. [30]	89.65	89.67	89.65	89.59
Dabass et al. [15]	97.70	97.69	97.73	97.71	Hage et al. [31]	99.53	99.33	99.33	99.33
Sharkas et al. [32]	99.20	99.24	99.23	99.24	Noaman et al. [33]	99.02	99.00	99.00	99.00
Proposed Method	99.40	99.38	99.47	99.42	Proposed Method	99.97	99.97	99.97	99.97
Model	LC25000-Colon				Model	LC25000-Lung & Colon			
	Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)		Acc. (%)	Pre. (%)	Rec. (%)	F1 (%)
Dabass et al. [34]	99.98	100.0	99.95	91.30	Provath et al. [14]*	99.76	99.60	99.40	99.70
Hage et al. [31]	99.30	99.50	99.50	99.50	Hage et al. [31]	99.00	98.60	99.00	98.80
Dabass et al. [15]	100.0	100.0	100.0	100.0	Zhu et al. [35]	97.52	97.52	97.53	97.51
Merabet et al. [36]	99.27	98.62	100.0	99.30	Uddin et al. [37]	99.96	99.96	99.96	99.96
Proposed Method	100.0	100.0	100.0	100.0	Proposed Method	99.96	99.96	99.96	99.96

Note: *Precision, recall, or F1 has been averaged class-wise for this cited article in Table IV. Bold indicates the results obtained by our model.

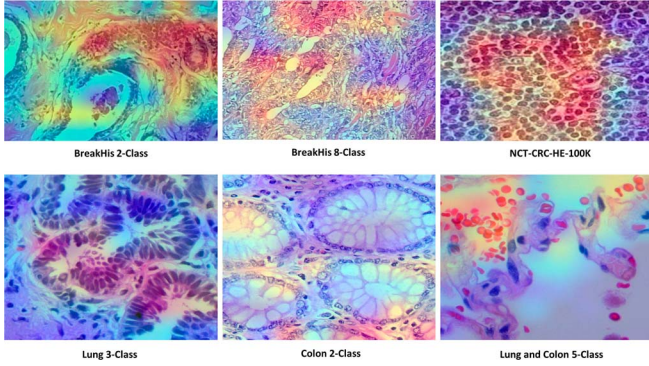


Fig. 10. GradCAMs of the SAWL-Net on each of the datasets.

Surface, Contour and Wireframe plot, to pictorially show some results obtained by the proposed model.

GradCAM Analysis: This study utilizes GradCAM, a method by Selvaraju et al. [38], which generates gradient-weighted class activation maps highlighting the region of interests that results in accurate classification as shown in Fig. 10 for the final Conv2D layer of the SAWL-Net.

t-SNE Plots: As per Van et al. [39], t-distributed Stochastic neighbor embedding (t-SNE) is a notable dimensionality reduction method for visualizing high-dimensional data in a lower-dimensional space. It transforms Euclidean distances between data points into conditional probability scores using Stochastic neighbor embedding (SNE). Fig. 11 illustrates the pixel-wise t-SNE plot on the datasets by our proposed model.

Activation Maps: Activation maps, derived from convolution operations in deep neural networks, represent 2-D grids showcasing local patterns in input images. These maps, displayed in Fig. 12, are crucial for visualizing and interpreting hierarchical

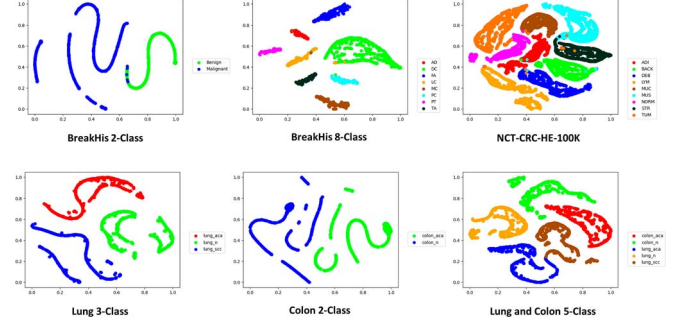


Fig. 11. t-SNE plots obtained by the SAWL-Net on each of the datasets.

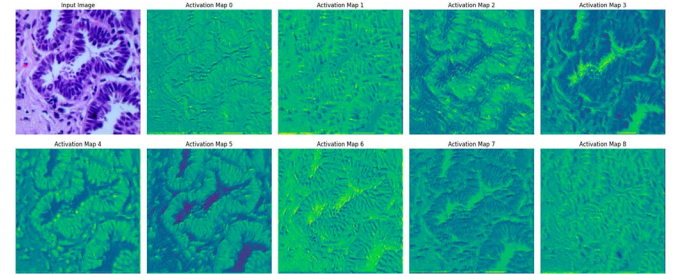


Fig. 12. Sample activation maps of the SAWL-Net model.

features learned by CNNs, providing insights into the network's comprehension of input data.

Surface, Contour and Wireframe Plots: These are used to visualize multidimensional data effectively. Surface plots show 3-D representations of functions or datasets, revealing behavior across input parameters. Contour plots display level curves, indicating consistent values within the dataset. Wireframe plots

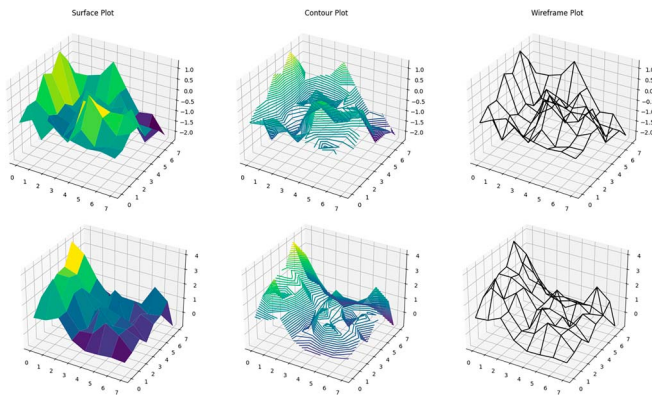


Fig. 13. Sample surface, contour and wireframe plots of the SAWL-Net model.

offer skeletal representations, highlighting underlying structures for thorough analysis. Employing these techniques on feature maps enhances understanding of gradient distributions, aiding decision-making and hypothesis testing as shown in Fig. 13.

V. CONCLUSION

We have introduced a novel histopathological image classification model, SAWL-Net, consisting of three key modules: the PCC module, the SRCC module and the wavelet transformation module. The PCC module leverages PCC and CS based channel attention mechanism based on linear and non-linear relationships, while SRCC uses SRC and CS based channel attention, while considering both ranked and directional relationships. The wavelet transformation module flattens channel-wise feature maps to wavelike format, while utilizing Conv1D layer for localized feature attention across all channel waves, providing a holistic approach. The integration of these components enhances the performance of the backbone, MobileNet, while keeping the model pretty lightweight, leading to SOTA results. Acknowledging the imperfections present in medical applications, we express the need for continuous improvements. Besides, we also plan to explore the model's performance across other medical image modalities.

REFERENCES

- [1] D. M. Metter, T. J. Colgan, S. T. Leung, C. F. Timmons, and J. Y. Park, "Trends in the US and Canadian pathologist workforces from 2007 to 2017," *JAMA Netw. Open*, vol. 2, no. 5, 2019, Art. no. e194337.
- [2] X. Wang et al., "Transpath: Transformer-based self-supervised learning for histopathological image classification," in *Proc. Medical Image Comput. Comp. Assisted Intervention (MICCAI): 24th Int. Conf.*, Strasbourg, France: Springer, Sep./Oct. 2021, pp. 186–195.
- [3] M. Khazaei Fadafan and K. Rezaei, "Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework," *Sci. Rep.*, vol. 13, no. 1, p. 8823, 2023.
- [4] C.-M. Lo, J.-K. Jiang, and C.-C. Lin, "Detecting microsatellite instability in colorectal cancer using transformer-based colonoscopy image classification and retrieval," *PLoS One*, vol. 19, no. 1, 2024, Art. no. e0292277.
- [5] J. C. Lee and J. T. Kwak, "Order-ViT: Order learning vision transformer for cancer classification in pathology images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 2493–2502.
- [6] S. B. Yengec-Tasdemir, Z. Aydin, E. Akay, S. Dogan, and B. Yilmaz, "An effective colorectal polyp classification for histopathological images based on supervised contrastive learning," *Comput. Biol. Med.*, vol. 172, 2024, Art. no. 108267.
- [7] C. Xu, K. Yi, N. Jiang, X. Li, M. Zhong, and Y. Zhang, "MDFF-Net: A multi-dimensional feature fusion network for breast histopathology image classification," *Comput. Biol. Med.*, vol. 165, 2023, Art. no. 107385.
- [8] G. Li et al., "Pathological image classification via embedded fusion mutual learning," *Biomed. Signal Process. Control*, vol. 79, 2023, Art. no. 104181.
- [9] S. Taheri, Z. Golrizkhatami, A. A. Basabrain, and M. S. Hazzazi, "A comprehensive study on classification of breast cancer histopathological images: Binary versus multi-category and magnification-specific versus magnification-independent," *IEEE Access*, vol. 12, pp. 50431–50443, 2024.
- [10] D. Addo et al., "A hybrid lightweight breast cancer classification framework using the histopathological images," *BioCybern. Biomed. Eng.*, vol. 44, no. 1, pp. 31–54, 2024.
- [11] C. Varun, A. K. R. Poreddy, T. Balakrishna, and P. Kokil, "A computer aided detection system for breast cancer using lightweight CNN models for smart healthcare," in *Proc. 10th Int. Conf. Bio Signals, Images, Instrum. (ICBSII)*, Piscataway, NJ, USA: IEEE Press, 2024, pp. 1–6.
- [12] M. Saha, M. Chakraborty, S. Maiti, and D. Das, "Breast-Net: a lightweight DCNN model for breast cancer detection and grading using histological samples," *Neural Comput. Appl.*, vol. 36, no. 32, pp. 20067–20087, 2024.
- [13] Z. Wang et al., "Dienet: an attention model for histopathological image grading of lung adenocarcinoma based on dimensional information embedding," *Sci. Rep.*, vol. 14, no. 1, p. 6209, 2024.
- [14] M. A.-M. Provath, K. Deb, P. K. Dhar, and T. Shimamura, "Classification of lung and colon cancer histopathological images using global context attention based convolutional neural network," *IEEE Access*, vol. 11, pp. 110164–110183, 2023.
- [15] M. Dabass, S. Vashisth, and R. Vig, "A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images," *Comput. Biol. Med.*, vol. 147, 2022, Art. no. 105680.
- [16] H. Chen et al., "IL-MCAM: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach," *Comput. Biol. Med.*, vol. 143, 2022, Art. no. 105265.
- [17] Z. Lv, Y. Lin, R. Yan, Y. Wang, and F. Zhang, "TransSurv: transformer-based survival analysis model integrating histopathological images and genomic data for colorectal cancer," *IEEE/ACM Trans. Computat. Biol. Bioinf.*, vol. 20, no. 6, pp. 3411–3420, Nov./Dec. 2023.
- [18] H. Yang, J.-Y. Kim, H. Kim, and S. P. Adhikari, "Guided soft attention network for classification of breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1306–1315, 2019.
- [19] H. Duanmu et al., "Spatial attention-based deep learning system for breast cancer pathological complete response prediction with serial histopathology images in multiple stains," in *Proc. Med. Image Comput. Comp. Assisted Intervention (MICCAI): 24th Int. Conf.*, Strasbourg, France: Springer, Sep./Oct. 2021, pp. 550–560.
- [20] 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.
- [21] H. N. M. A. Kather, J. N., 100,000 Histological Images of Human Colorectal Cancer and Healthy Tissue (v0.1) [Data Set], Organization: Zenodo, 2018, doi: 10.5281/zenodo.1214456.
- [22] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lc25000 lung and colon histopathological image dataset," 2019, *arXiv:1912.12142v1*.
- [23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [24] S. Roy, P. K. Jain, K. Tadepalli, and B. P. Reddy, "Forward attention-based deep network for classification of breast histopathology image," *Multimedia Tools Appl.*, vol. 83, pp. 88039–88068, Dec. 2024.
- [25] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of xgboost-based algorithm for breast cancer classification," *Mach. Learn. Appl.*, vol. 6, 2021, Art. no. 100154.
- [26] C. S. Vikranth, B. Jagadeesh, K. Rakesh, D. Mohammad, S. Krishna, and R. A. As, "Computer assisted diagnosis of breast cancer using histopathology images and convolutional neural networks," in *Proc. 2nd InterNat. Conf. Artif. Intell. Signal Process. (AISP)*, Piscataway, NJ, USA: IEEE Press, 2022, pp. 1–6.

- [27] P. Jeevan and A. Sethi, "Which backbone to use: a resource-efficient domain specific comparison for computer vision," 2024, *arXiv:2406.05612*.
- [28] A. Halder and D. Dey, "MorphattnNet: an attention-based morphology framework for lung cancer subtype classification," *Biomed. Signal Process. Control*, vol. 86, 2023, Art. no. 105149.
- [29] A. Kumar, A. Vishwakarma, and V. Bajaj, "CRCCN-Net: Automated framework for classification of colorectal tissue using histopathological images," *Biomed. Signal Process. Control*, vol. 79, 2023, Art. no. 104172.
- [30] R. Fan et al., "One-vote veto: Semi-supervised learning for low-shot glaucoma diagnosis," *IEEE Trans. Med. Imag.*, vol. 42, no. 12, pp. 3764–3778, Dec. 2023.
- [31] A. Hage Chehade, N. Abdallah, J.-M. Marion, M. Oueidat, and P. Chauvet, "Lung and colon cancer classification using medical imaging: A feature engineering approach," *Phys. Eng. Sci. Med.*, vol. 45, no. 3, pp. 729–746, 2022.
- [32] M. Sharkas and O. Attallah, "Color-CADX: a deep learning approach for colorectal cancer classification through triple convolutional neural networks and discrete cosine transform," *Sci. Rep.*, vol. 14, no. 1, p. 6914, 2024.
- [33] N. F. Noaman, B. M. Kanber, A. A. Smadi, L. Jiao, and M. K. Alsmadi, "Advancing oncology diagnostics: AI-enabled early detection of lung cancer through hybrid histological image analysis," *IEEE Access*, 2024.
- [34] M. Dabass, J. Dabass, S. Vashisth, and R. Vig, "A hybrid u-net model with attention and advanced convolutional learning modules for simultaneous gland segmentation and cancer grade prediction in colorectal histopathological images," *Intell.-Based Med.*, vol. 7, 2023, Art. no. 100094.
- [35] H. Zhu, W. Liu, Z. Gao, and H. Zhang, "Explainable classification of benign-malignant pulmonary nodules with neural networks and information bottleneck," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 2, pp. 2028–2039, Feb. 2025.
- [36] A. Merabet, A. Saighi, M. A. Ferradji, and Z. Laboudi, "Enhancing colon cancer prediction in histopathology with integrated deep learning models: A comparative study on the LC25000 dataset," in *Proc. 6th Int. Conf. Pattern Anal. Intell. Syst. (PAIS)*, Piscataway, NJ, USA: IEEE Press, 2024, pp. 1–7.
- [37] A. H. Uddin, Y.-L. Chen, M. R. Akter, C. S. Ku, J. Yang, and L. Y. Por, "Colon and lung cancer classification from multi-modal images using resilient and efficient neural network architectures," *Heliyon*, vol. 10, no. 9, 2024.
- [38] 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.*, 2017, pp. 618–626.
- [39] L. Van der Maaten and G. Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.