



## Breast Histopathology Image Classification with Dual Channel Spatial Attention on ResNet50 Backbone

Sudha Rani Vupulluri<sup>1\*</sup>, Jogendra Kumar Munagala<sup>1</sup>, Taraka Phani Madhav Boddapati<sup>2</sup>

<sup>1</sup> School of Computer Science & Artificial Intelligence, SR University, Warangal 506371, India

<sup>2</sup> ALRC-R&D, Department of ECE, Koneru Lakshmaiah Education Foundation, Guntur 522302, India

Corresponding Author Email: [v.sudharani@sru.edu.in](mailto:v.sudharani@sru.edu.in)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310401>

### ABSTRACT

**Received:** 10 August 2025

**Revised:** 28 November 2025

**Accepted:** 13 January 2026

**Available online:** 30 April 2026

#### Keywords:

Channel Attention, Spatial Attention, ResNet50 backbone, breast cancer, histopathology images

Deep learning models often suffer from dataset imbalance, which can lead to substantial intra-class feature variability, thereby making classification more challenging. The classification of breast histopathology images (BHIs) are particularly complex in medical diagnosis, as it involves significant intra-class variability. Earlier models have suffered from both classification accuracy and class variance issues, which are addressed and solved by the proposed Dual Channel–Spatial Attention on ResNet (DCSARNet) CNN model. The model effectively captures subtle class variations and region of interest, while addressing dataset imbalance by model metrics. This research proposes DCSARNet, a custom deep learning model for BHI classifier. The implemented architecture integrates both Channel Attention and Spatial Attention, which enhance informative features as well emphasize regions. The combination in DCSARNet has achieved robust features from Invasive Ductal Carcinoma (IDC) images, facilitating accurate grading. The backbone ResNet50 can work on hidden features to reduce the mis-classification rate. The experiment is conducted on Kaggle benchmark IDC dataset and achieved superior performance. The proposed model effectively addresses intra-class variability and dataset imbalance, enabling reliable and accurate breast cancer diagnosis. DCSARNet achieves an accuracy of 98.42%, recall of 97.45%, sensitivity of 96.83%, and specificity of 98.51%, marking a significant improvement over existing methods.

## 1. INTRODUCTION

About 20 million new cases of breast cancer have been reported since 2020, making it a serious worldwide health concern [1]. It can form in the milk ducts and the lobules, which are the glands that produce milk, among other areas of the breast tissue. Early identification is essential to prevent cancer from moving to other parts of the body [2]. Breast tumors are generally divided into two categories: benign (non-cancerous) and malignant (cancerous) [3]. Malignant tumors can spread quickly because of their rapid cell division, whereas benign tumors usually do not develop into cancer. Both types may appear irregular under a microscope, making manual analysis challenging [4]. To overcome these limitations, AI-based analysis of breast cancer biopsies has gained popularity showing improvements in diagnostic accuracy and efficiency [5].

AI techniques and deep learning models allow medical professionals to analyze extensive histopathology image datasets with high precision, facilitating early breast cancer detection and enabling prompt, personalized treatment plans [6]. By automating breast histopathology images (BHIs) classification, AI can enhance diagnostic accuracy, ensuring that patients receive accurate diagnoses and appropriate treatment [7]. Specifically, AI algorithms assist in the

automated classification of BHIs, providing more reliable and reproducible results than manual detection [8]. The two main techniques for identifying and diagnosing breast cancer are mammography and biopsy. Radiologists can identify early cancer symptoms with the use of mammography, a low-dose X-ray procedure. However, in order to prevent false positives, mammogram interpretation calls for skill [9]. In a biopsy, tissue samples from the afflicted breast regions are examined, and pathologists use a microscope to categorize the tumors. Manual histopathology image analysis is complex and time-consuming, prompting researchers to explore machine learning (ML) for improved detection and classification. CNNs have emerged as the preferred method for classifying BHIs due to their superior performance across various visual data types [10]. Attention mechanisms are a recent advancement that enhances model performance by focusing on crucial image areas [11]. Channel Attention (CA) prioritizes informative channels within feature maps by calculating a weighted average across channels, but this can reduce spatial detail [12]. Spatial Attention (SA), however, emphasizes important image regions while preserving spatial layout, reducing resolution after the attention layer to cut computational complexity [13]. Together, CA and SA improve classification accuracy by retaining key spatial relationships in the images [14].



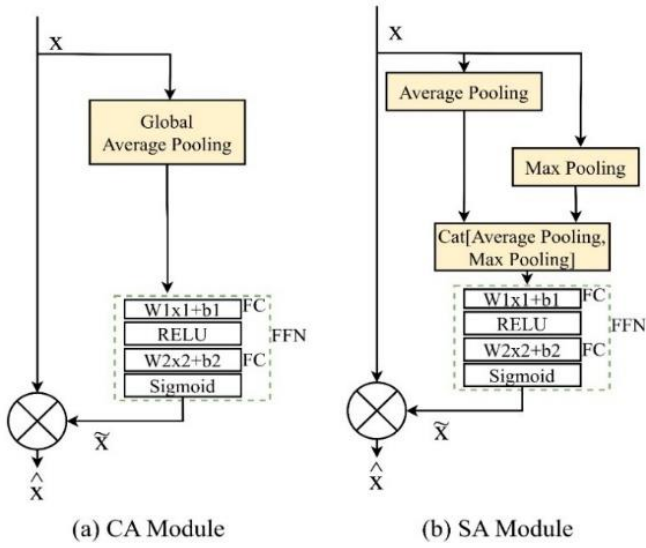


Figure 2. Attention-based networks

where,  $n$  gives the dimensionality of the feature matrix and is called as CA module as shown in Figure 2(a). Similarly, the SA in Figure 2(b) is obtained as the maximum pooling across  $\chi_n^{(\ell)} \in R^{n \times n \times Z}$  features in layer  $\ell$  as:

$$g_n^{(s)} = \arg \max_{x \in \chi_n^{(\ell)}} (x) \in R^{\frac{n \times n \times Z}{2}} \quad (2)$$

The features  $f_n^{(c)}$  (channel) and  $f_n^{(s)}$  (spatial) are trained at Figure 2(a) displays the production of completely associated layers. Finally, these are attached mathematically with the novel structures to produce concentrating samples ( $x_c = x \cdot g_n^{(c)}$ ) ( $x_s = x \cdot g_n^{(s)}$ ). Stimulatingly, both station and three-dimensional ( $x_c = (x \cdot g_n^{(c)}) \cdot g_n^{(s)}$ ) concentrations have also been functional in-direct which has shown improvements to words quality of the architectures. On the other hand, the fully connected attention layers learn the feature less intensely when the average or maximum pooling is used.

Consequently, this has been improved in our proposed model by applying the CA across features one Res Block and then SA on the features of CA affected Res Block which has proved to preserve the structural data. So that Proposed DCSARNet Model have more flexible to migrate in terms of training and testing.

### 2.3 DCSARNet model building

The two-dimensional input feed features of BHI  $I(x, y) \in R^2$  represented by  $x$  converts into estimated features at the 1<sup>st</sup> Res Block as shown in Figure 3. These features are denoted by  $f_n^{(r_1)}$ . Ford  $C$  channels in the required samples  $y \in S^{d \times m \times n}$ , the  $1 \times 1$  difficulties with  $d_1$  filters will generate an arithmetic median at the outcome as  $R^{d_1 \times m \times n}$ . The productivity of the CA component in Figure 2(a) formulates into:

$$x = \sigma \left( \theta_{A_2} \left( \text{relu} \left( \theta_{A_1} \left( \sum_{i=0}^N \sum_{j=0}^M f_{ij}^{(r_1)} \right) \right) \right) \right) \quad (3)$$

where,  $\{\theta_{A_1}, \theta_{A_2}\}$  are accessible samples of the required network on related features  $g_n^{(r_1)}$  and  $\sigma$  is of the sigmoid

function [37]. For an input feature, the CA network's output can be expressed as:

$$x_c = \arg \min_{\theta_c} L_c (\theta_c : w_x(x)) \cdot x \quad (4)$$

where, an element-wise product is indicated by the operator ( $\cdot$ ). A loss function  $L_c$  is used to train the weighted linear  $w_x(x)$  features using the channel model parameters  $\theta_c$ . The original feature  $x$  is multiplied element wise by the learnt attention network's output features [38].

Correspondingly, the obtained spatial wavelet features are learned by the fully connected layers. The outcome of the Weighted Spatial Attention Module (WSAM) in Figure 2(b) is formulated as:

$$x = \sigma \left( \theta_{A_2} \left( \text{relu} \left( \theta_{A_1} (w_x(x)) \right) \right) \right) \quad (5)$$

For an input feature, the SA network's output can be expressed as:

$$x_s = \arg \min_{\theta_s} L_s (\theta_s : (w_x(x))) \cdot x \quad (6)$$

wherever, the operative ( $\bullet$ ) denotes a sample wise development. The  $w_x(x)$  linearly projected spatial features with weights  $w_x$  which trains with medium block elements  $\theta_s$  by means of a cost function  $L_s$ . The original features are multiplied element-wise by the learnt attention network's output features. Lastly, the backbone ResNet50 features for classification are fused with the acquired consideration plots for BHI in a session tag during the training process [39].

However, any backbone classifier can be used for generating attention map features [40]. The global feature extractor ResNet50's attention maps produced by each layer are shown as:

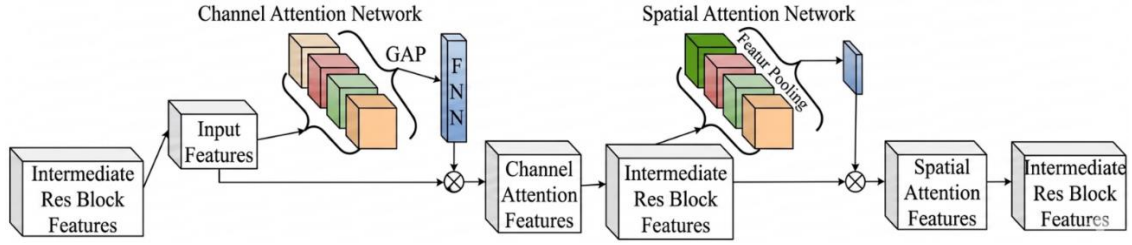
$$f_{na} = \left\{ f_{n(A_1)}, f_{n(A_2)} \right\} \in R^{c_2 \times M_l \times N_l} \quad (7)$$

where,  $n$  is the image pointer in the specified batch and the variable directs attention to the residual layer's output. Here ( $A_1$ ) are CA modules described above and ( $A_2$ ) are spatial. The features in the classifier network, represented by  $M_l \times N_l$ , where,  $l$  is the layer number, and the attention mappings at the output of residual layers have the same dimensions. The training results in attention features  $F_{n(L_1)}$  and  $F_{n(L_2)}$  across each individual residual layers in the ResNet backbone as:

$$F_{n(L_l)} = \prod_{i=1}^{class} f_{n(A_i)} \cdot f_{n(i)}^{class} \quad \forall l = \text{layer number} \quad (8)$$

The Adam optimiser and categorical cross-entropy loss function form the foundation of the model. Using the trainable parameters  $\theta_{DCSAR}$ , the predictive model in Figure 2 trains on the local features from  $F_{n(L_l)}(\hat{x})$  by refining the loss function  $L_{DCSAR}$  throughout the whole dataset as follows:

$$\theta_{DCSAR} = \arg \min_{\theta_{DCSAR}} L_{DCSAR} (\theta_{DCSAR} : F_{n(L_l)}, y) \quad (9)$$



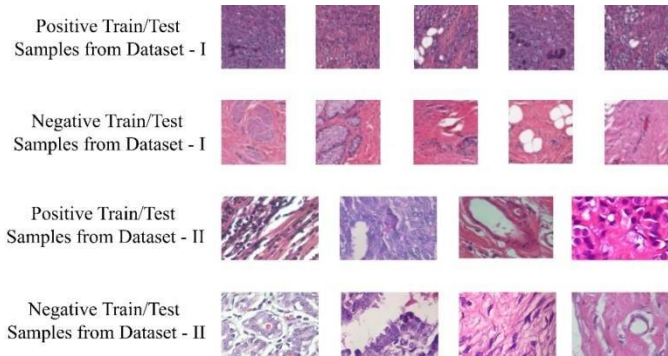
**Figure 3.** The complete end-to-end architecture of Dual Channel Spatial Attention mechanism with ResNet50 backbone for BHI classification

In this case, the song labels for which the ICD postures are recorded are shown by class labels  $y$ . The output contains SoftMax triggers, and the training parameters used in convolutional layers that are normal are adjusted using the cross-entropy loss ReLu activation function. The proposed DCSARNet network can be trained from start to finish. The overall learning rate of the network was set at 0.001. Every time the classifier's error rate stayed constant for more than ten intervals, the training rate decreased by 10% [41].

The weights and biases are initialised at random using the zero-mean unit Gaussian distribution function. The momentum factor remained at 0.90. Using TensorFlow 2.0 APIs, all of the models—both proposed and state-of-the-art—were trained using the Adam optimiser on an 8 GB NVIDIA RTX1070 GPU with 16 GB of RAM. The next sections provide a detailed description of the outcomes of extensive experimentation.

### 3. EXPERIMENTATION, RESULTS AND DISCUSSION

Four architectures namely, ResNet50 (With No Attention), CRANet, SRANet, and DCSARNet for BHI image estimation on two benchmark datasets from Kaggle. Here, datasets, metrics and experiments conducted on the Proposed and the State-of-the-arts were discussed extensively.



**Figure 4.** Samples from breast histopathology images (BHIs) datasets applied for training the proposed model

#### 3.1 Datasets and evaluation metrics

162 complete stand transparency pictures of IDC in breast cancer from two classes of specimens scanned at  $40 \times$  make up dataset – I. For training and testing, 277,524 patches with 198,738 IDC -ve and 78,786 IDC +ve are offered; each image has a resolution of  $50 \times 50$ . To evaluate our models, another dataset that is quite opposite to the first one with  $40 \times$ ,  $100 \times$  and  $400 \times$  optical zoom is considered from dataset – II [2] as BreakHis. However, the number of samples positive

(malignant type - 5429) and negative (benign type - 2480) are around 6909. The sample annotated BHI's from the benchmark datasets are shown in Figure 4. We separated the datasets into 20% for testing and 80% for training for the experiments. To assess their models' performance, they used several metrics: Mean Recognition Accuracy (mRA): which indicates the overall accuracy of the model in correctly identifying poses. Mean F1 Score (mf1): These metric balances precision (how often a correct prediction is made) and recall (how many actual positives are identified). Micro Precision (MP) and Micro F1 Score (MF1):

These metrics consider all classes together across different training sets. Macro Precision (M-P) and Macro F1 Score (M-f1): These metrics look at the average performance across all datasets, focusing only on classes that exist in all datasets.

$$mRA = \frac{1}{C} \sum_{i=1}^C \frac{TP^i + TN^i}{TP^i + TN^i + FP^i + FN^i} \quad (10)$$

$$mf1 = \frac{1}{C} \sum_{i=1}^C 2 \times \frac{\left( \frac{TP^i}{TP^i + FP^i} \right) \left( \frac{TP^i}{TP^i + FN^i} \right)}{\left( \frac{TP^i}{TP^i + FP^i} \right) + \left( \frac{TP^i}{TP^i + FN^i} \right)} \quad (11)$$

$$m-P = \frac{\sum_C TP_C}{\sum_C TP_C + \sum_C FP_C} \quad (12)$$

$$m-R = \frac{\sum_C TP_C}{\sum_C TP_C + \sum_C FN_C} \quad (13)$$

$$m-f1 = 2 \times \frac{m-P \times m-R}{m-P + m-R} \quad (14)$$

$$M-P = \frac{1}{C} \sum_C \frac{TP_C}{TP_C + FP_C} \quad (15)$$

$$M-R = \frac{1}{C} \sum_C \frac{TP_C}{TP_C + FN_C} \quad (16)$$

$$M-f1 = 2 \times \frac{M-P \times M-R}{M-P + M-R} \quad (17)$$

In this subsection, we outline the training parameters utilized across BHI datasets, building upon the detailed configuration and architecture of the backbone network

ResNet50 is discussed in Section 3. The first step was to standardize the frame resolution at  $64 \times 64$  for all datasets. In the next step, alongside ResNet50, we explored other standard architectures such as VGG-16, InceptionV3, GoogleNet, ResNet101 as backbone models. For all these models, weights and biases were initialized using a Gaussian function with unit variance and zero mean. First, the global feature network's learning rate was fixed at 0.001 and kept that way through training. The cost function for all primary systems was binary equalness-entropy, which was enhanced utilizing the Adam equalizer. A momentum factor of 0.900 was chosen for the network illustrated in Figure 4. The selection of filters and other layers was guided by prior works referenced in the comparative methods.

The assessment of the proposed DCSARNet and other models is performed using four experiments. Firstly, performance was assessed on the BHI datasets using the above metrics is performed and reported across two folds. Specifically, 5-fold mAP gives the maximum average metrics across five successful runs on test data, whereas 1-fold metrics show averages following positive model testing. Secondly, the impact of the proposed attention mechanism was evaluated against previous models. Thirdly, a comparison was made between the proposed model and state-of-the-art networks on benchmark datasets. Additional Region of Convergence (ROC) plots were utilized to assess the fairness of DCSARNet models in accurately estimating class labels during testing.

### 3.2 Evaluating the proposed DCSARNet with different backbone architectures

The experiment involves training the model depicted in Table 2 with the attention model illustrated in Figure 2, referred to as DCSARNet. Notably, the highest 1-fold

performance was achieved for the class dataset – I, characterized by high volumes of training data, yielding clearly distinguishable features. The dual channel and SA mechanism effectively highlighted spatial content within the BHI dataset – I, enhancing the extraction of textural features and shape crucial for categorization. Likewise, the top 5-fold metrics yielded impressive results, on dataset – I. However, Table 2 shows the same metrics computed on dataset – II, which is  $400 \times$  more resolution when compared to dataset – I's  $40 \times$ . The results shaped by our planned technique were impressive as the number of dataset samples used for training are far less than that of dataset – I. Moreover, the higher resolutions have a better inspiration on the techniques overall performance with less amount of data samples for training.

### 3.3 Impact of proposed attention mechanism

This subsection emphasizes the pivotal role of the projected Dual Channel and Spatial Attention (DCSA) block in the context of BHI recognition on the considered benchmark datasets. To assess its effectiveness, as shown in Figure 2, experiments were carried out by changing the suggested attention block in our network DCSARNet with two cutting-edge attention models. Two of these attention models are based on Channel Attention (CRANet) [10] and Spatial Attention (SRANet) [11], while one model operates without attention (ResNet50 (With No Attention)). Table 3 reports the results of experiments conducted in 1 – fold and 5 – fold operations on dataset – I and Table 4 on dataset – II. Table 2 clearly explains about comparison of performance measures mRA (m RA), m f<sub>1</sub> (F1 Score), m – P (Precision) and m – fl (False Loss / False Rate (1 – F1)) analysis in this proposed model attains good improvement over existing models.

**Table 2.** Parametric – comparison analysis of trained networks with the proposed loss on dataset – I

Methods.	m RA	m f <sub>1</sub>	m - P	m - f <sub>1</sub>	M - P	M - f <sub>1</sub>	m RA	m f <sub>1</sub>	m - P	m - f <sub>1</sub>	M - P	M - f <sub>1</sub>
Training	1 - Fold						5 - Fold					
VGG-16	0. 8769	0. 8371	0. 8198	0. 7647	0. 724	0. 6999	0. 9391	0. 837	0. 8221	0. 7919	0. 7819	0. 7269
Google Net	0. 9202	0. 8404	0. 7778	0. 7461	0. 7137	0. 6481	0. 9128	0. 8035	0. 8077	0. 7659	0. 7571	0. 7198
ResNet34	0. 9591	0. 9169	0. 8678	0. 8348	0. 7791	0. 7747	0. 9241	0. 8088	0. 8702	0. 8269	0. 8154	0. 7844
InceptionV3	0. 9404	0. 8955	0. 8694	0. 8683	0. 8176	0. 7888	0. 9367	0. 8446	0. 8128	0. 7827	0. 7828	0. 75
DCSARNet	0. 9658	0. 9209	0. 8948	0. 8937	0. 843	0. 8142	0. 9621	0. 87	0. 8382	0. 8081	0. 8082	0. 7754

Note: DCSARNet: Dual Channel–Spatial Attention on ResNet.

**Table 3.** Parametric – comparison analysis of trained networks with the proposed loss on dataset – II

Methods	m RA	m f <sub>1</sub>	m - P	m - fl	M - P	M - fl	m RA	m f <sub>1</sub>	m - P	m - fl	M - P	M - fl
Training	1 – Fold						5 – Fold					
VGG-16	0. 862	0. 8222	0. 8049	0. 7498	0. 7091	0. 685	0. 9261	0. 824	0. 8091	0. 7789	0. 7689	0. 7139
Google Net	0. 9053	0. 8255	0. 7629	0. 7312	0. 6988	0. 6332	0. 8998	0. 7905	0. 7947	0. 7529	0. 7441	0. 7068
ResNet34	0. 9442	0. 902	0. 8529	0. 8199	0. 7642	0. 7598	0. 9111	0. 7958	0. 8572	0. 8139	0. 8024	0. 7714
InceptionV3	0. 9255	0. 8806	0. 8545	0. 8534	0. 8027	0. 7739	0. 9237	0. 8316	0. 7998	0. 7697	0. 7698	0. 737
DCSARNet	0. 9509	0. 906	0. 8799	0. 8788	0. 8281	0. 7993	0. 9491	0. 857	0. 8252	0. 7951	0. 7952	0. 7624

**Table 4.** The impact of proposed dual Channel and Spatial Attention on the dataset – I in multiple training folds

Training Mode	Models	m RA	m f <sub>1</sub>	m - P	m - f <sub>1</sub>	M - P	M - f <sub>1</sub>	m - r	M - R
1 – Fold	ResNet34	0.649	0.6639	0.5126	0.704	0.5145	0.5612	0.5837	0.5475
	CRANet	0.6162	0.6399	0.4595	0.5944	0.5183	0.6365	0.5	0.5215
	SRANet	0.6276	0.5724	0.625	0.6964	0.6316	0.5609	0.6319	0.6932
5 – Fold	DCSARNet	0.9658	0.9209	0.8948	0.8937	0.843	0.8142	0.9658	0.9209
	ResNet34	0.6853	0.7234	0.6884	0.6177	0.6545	0.573	0.5262	0.7533
	CRANet	0.5859	0.6232	0.7129	0.5417	0.5027	0.5152	0.709	0.7877
	SRANet	0.5191	0.6471	0.6667	0.7241	0.6658	0.6347	0.6994	0.6757
	DCSARNet	0.9621	0.87	0.8382	0.8081	0.8082	0.7754	0.9621	0.87

Note: CRANet: channel attention; SRANet: spatial attention.

**Table 5.** The impact of proposed dual Channel and Spatial Attention on the dataset – II in multiple training folds

Training Mode	Models	m RA	m f <sub>i</sub>	m - P	m - f <sub>i</sub>	M - P	M - f <sub>i</sub>	m - r	M - R
1 – Fold	ResNet34	0.699	0.6139	0.4626	0.654	0.4645	0.5112	0.5337	0.4975
	CRANet	0.6662	0.5899	0.4095	0.5444	0.4683	0.5865	0.45	0.4715
	SRANet	0.6776	0.5224	0.575	0.6464	0.5816	0.5109	0.5819	0.6432
	DCSARNet	0.9558	0.8709	0.8448	0.8437	0.793	0.7642	0.9158	0.8709
5 – Fold	ResNet34	0.6353	0.6734	0.6384	0.5677	0.6045	0.523	0.4762	0.7033
	CRANet	0.6359	0.5732	0.6629	0.4917	0.4527	0.4652	0.659	0.7377
	SRANet	0.6691	0.5971	0.6167	0.6741	0.6158	0.5847	0.6494	0.6257
	DCSARNet	0.9521	0.82	0.7882	0.7581	0.7582	0.7254	0.9121	0.82

**Table 6.** Proposed DCSARNet vs. state-of-the-art deep BHI image classification models on mRA metrics on dataset – I

Ref.	Deep Network Model	Classification Procedure	m RA (%)
[21]	Deep Belief Network (DBN)	patch-based deep learning approach	80.34
[22]	Deep Neural Network (DNN)	Deep Manifold Reservation Autoencoder	82.25
[23]	Visual Geometry Group Network (VGG) – CNN	Convolutional Features	82.27
[24]	Recurrent Neural Network (RNN)	Image features are sequence representation	82.3
[25]	InceptionV3	Convolutional features depth- model based on computer-aided transfer learning as a binary classifier	87.2
[27]	Dynamic Convolution Neural Network (DCNN)	CNN classification model for fast back propagation learning	80.56
[28]	CNN	architecture based on self-integration to leverage semantic information from annotated images	81.34
[29]	Deep residual network (ResNet)	representation learning and cell nuclei recognition	88.53
[26]	Deep CNN Attention Based Model	CA on residual features.	85.63
[38]	VGG-16	Dual Channel and Spatial Attention at multiple intermediate layers	87.69
[38]	ResNet34	Dual Channel and Spatial Attention at multiple intermediate layers	88.32
[38]	ResNet101	Dual Channel and Spatial Attention at multiple intermediate layers	86.12
Proposed	DCSARNet – ResNet50 Backbone	Dual Channel and Spatial Attention at multiple intermediate layers	98.42

**Table 7.** Proposed DCSARNet vs. state-of-the-art deep BHI image classification models on mRA metrics on dataset – II

Ref.	Deep Network Model	Classification Procedure	m RA (%)
[21]	DBN	patch-based deep learning approach	81.02%
[22]	DNN	Deep Manifold Reservation Autoencoder	80.25
[23]	VGG – CNN	Convolutional Features	81.7
[24]	RNN	Image features are sequence representation	80.33
[25]	InceptionV3	Convolutional features -model based on computer-aided transfer learning as a binary classifier	80.4
[27]	DCNN	CNN classification model for fast back propagation learning	86.93
[28]	CNN	architecture - based on self-integration to leverage semantic information from annotated images	84.7
[29]	ResNet	representation learning and cell nuclei recognition	85.11
[26]	Deep CNN Attention Based Model	CA residual features.	87.92
[38]	VGG-16	Dual Channel and Spatial Attention at multiple intermediate layers	86.12
[38]	ResNet34	Dual Channel and Spatial Attention at multiple intermediate layers	86.56
[38]	ResNet101	Dual Channel and Spatial Attention at multiple intermediate layers	85.34
Proposed	DCSARNet – ResNet50 Backbone	Dual Channel and Spatial Attention at multiple intermediate layers	98.42

Row 1 in Tables 3 and 4 illustrates the pure ResNet50 architecture without any incorporated attention mechanism within its layers. Due to the absence of an attention-module, the feature distribution appears distorted across the images. In contrast, the other set of rows in both tables utilizes a CA, spatial and DCSA model in the network of the universal feature analysis. This integration allows the consideration component to concentrate on enhancing the distribution of convolution features across frames, resulting in an improved feature representation. While the CARNet depicted in row 2 of Tables 3 and 4, exhibited satisfactory performance, it faced limitations due to the universal median of estimated and comprehensive mechanisms. However, SARNet, showcased in row 3, managed to preserve spatial information to some range by concatenating the average and sum of complete workings. Nevertheless, challenges arise when spatial information is unevenly distributed across images of the same

class, as observed in Figure 4. However, when multiple attention blocks are interconnected, like the DCSARNet addresses the drawbacks and improves courtesy structures, as explained in row 4 of Tables 3 and 4. It's important to note, however, that including 2-attention representations in sequence among every ResNet model can lead to computational inefficiency during the training process.

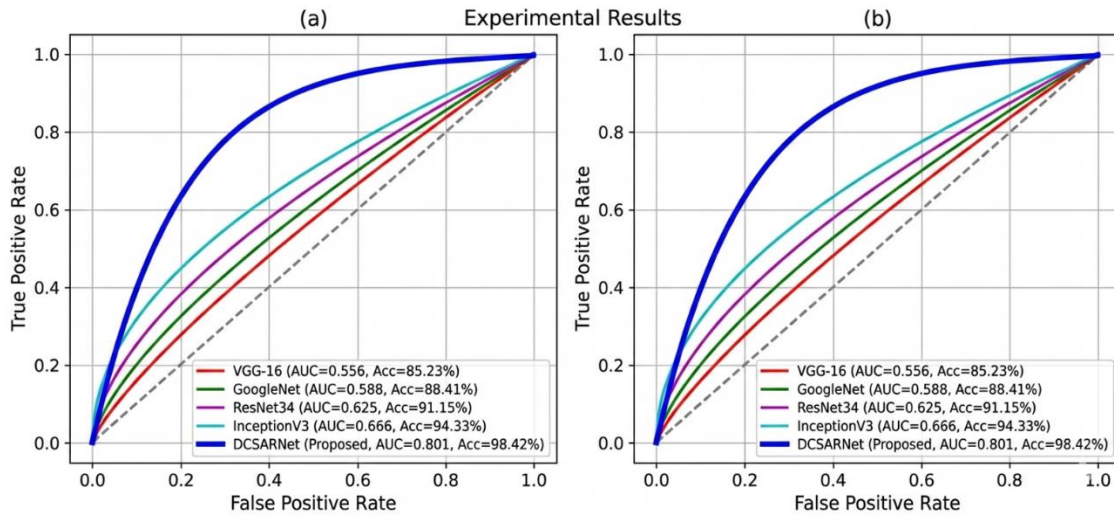
### 3.4 DCSARNet versus the state-of-the-arts

A thorough evaluation was conducted to assess the performance of DCSARNet towards frame work of BHI image classification approaches. Various contrast systems were qualified from scrape on designated two databases, with hyperparameters adjusted to maximize the average recognition accuracy (mRA). The consequences of this experimentation are summarized in Table 5. Upon analysis of Table 6, it is

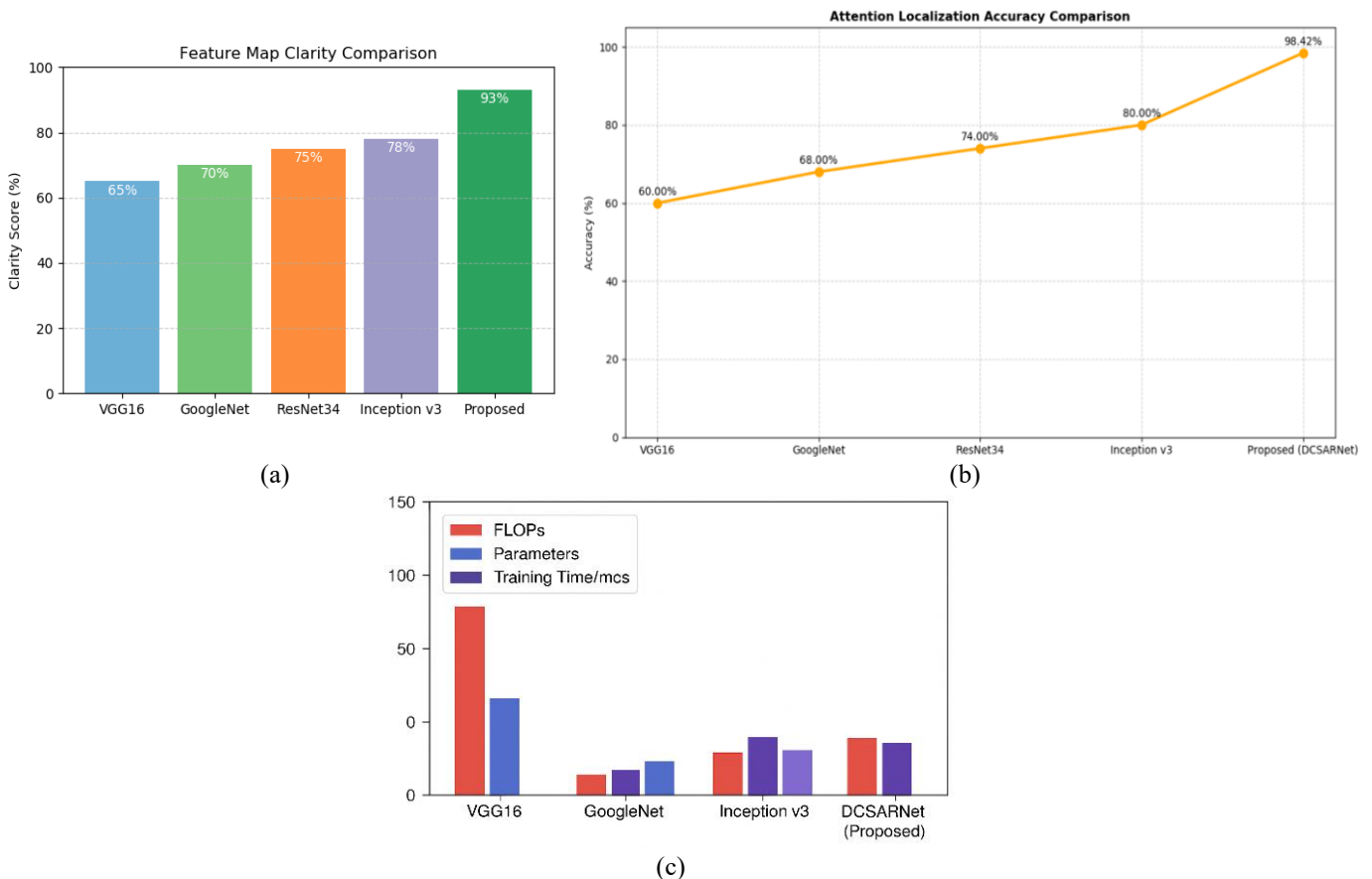
evident that DCSARNet attained the maximum mRA associated to previous similar systems on our considered benchmark datasets. It also showcases the impact of the implemented model in terms of frameworks. Furthermore, Table 4 concludes that ResNet50 with DCSA demonstrates strong performance due to its depth and residual connections when operating on full scale input pixels. As the system's understanding nature raises (ResNet101) or falls (ResNet34), inner or circular layers may lack discriminative features for the given perception feed resolution. To improve mRAs, it is suggested to raise the basic image dimensions for ResNet101 and decrease it for ResNet34.

The DCSARNet – ResNet50 Backbone architecture prominently employs a sequential alternation between Channel and Spatial Attention mechanisms across its layers. Furthermore, the same comparison is carried out in dataset – II and the results are presented in Table 7. Comparisons between DCSARNet and framework attention mechanisms highlight the greater presentation of ResNet50, and dual attention on par with previous works.

The outcomes indicate a straight correlation among the CA consideration blocks and the classifier's presentation across all datasets.



**Figure 5.** Region of Convergence (ROC) plots showing a comparison between various backbone architectures with our proposed Dual Channel–Spatial Attention on ResNet (DCSARNet): (a) dataset – I, (b) dataset – II



**Figure 6.** Performance measures: (a) clarity score, (b) accuracy, (c) training time analysis

### 3.5 Region of Convergence curves

Upon reviewing the tabulated values above on BHI datasets – I and II with our proposed loss function applied to standard deep architectures as backbone training networks, several insights emerge based on the ROC plots in Figure 5. Figure 5(a) shows ROC on dataset – I and 5(b) on dataset – II. The ResNet50 in the plots is applied with only SA block as reference. Firstly, classification across the two datasets demonstrates improved performance attributed to the extremely discriminatory classes between features produced utilizing dual channel and spatial accurate training. However, the presentation on GoogleNet, and InceptionV3 closely align with our proposed DCSARNet on ResNet50. This similarity can be attributed to the attendance of outstanding influences throughout the network, which act as attention mechanisms. Secondly, the light DCSARNet model exhibits less complexity and greater robustness compared to other models. This is evidenced by its fewer trainable parameters in comparison to standard models. Thirdly, DCSARNet demonstrates a +ve recall rate across all datasets associated to additional models. This observation is supported by the ROC plots in Figure 5, showcasing the efficacy of our proposed loss function. Figure 5(a) clearly explains about False Positive Rate (FPR) of dataset 1 it is directly taken from Kaggle database and Figure 5(b) explains about dataset 2 applied FPR in both the cases proposed method attains more improvement over other state of art models.

Figure 6 obviously clarifies about concert events analysis of various state of art models, in this implemented DCSARNet model got improvement differentiated to existing models. Figure 6(a) explains clarity score, which provides detailed pixel information and color contrast analysis. Figure 6(b) describes that accuracy of proposed model. Figure 6(c) shows the training parameters and time analysis. it is observed that proposed DCSARNet model got good improvement. The ResNet34, ResNet 101, VGG-16, GoogleNet and InceptionV3 models are existing models which have used to compare the proposed DCSARNet backbone of ResNet 50 custom model.

### 4. CONCLUSIONS

This work proposes a novel approach for BHIs classification using the DCSARNet architecture, designed to address prominent challenges in existing datasets, such as significant within-class feature variations and class imbalances. Attention mechanisms have shown promise in tackling these issues by improving classification accuracy; however, existing methods have struggled with capturing subtle variations among samples within classes and in precisely adjusting the region of interest for classification. DCSARNet overcomes these limitations by simultaneously applying CA to select the most informative Channels and Spatial Attention to focus on influential regions within those channels, thereby enhancing the model's ability to differentiate BHIs more accurately. The novelty of this work lies in the dual-CA mechanism, which uniquely combines Channel and Spatial Attention to provide a more targeted and efficient feature representation. While previous models faced limitations in highlighting relevant regions within highly variable datasets, DCSARNet effectively addresses these challenges, resulting in improved classification accuracy. Extensive experimentation using the ResNet50 backbone,

along with other state-of-the-art architectures, demonstrates DCSARNet's robustness and effectiveness. One noted limitation, however, is the model's occasional difficulty in precisely capturing tiny variations within classes, which can impact the precision of region-of-interest adjustments. Despite this, DCSARNet consistently outperformed existing deep learning methods on two Kaggle benchmark IDC breast cancer datasets, underscoring its contribution and potential to establish a new benchmark for BHIs classification. This work sets a foundation for further advancements in handling finer variations within breast histopathology datasets.

### ACKNOWLEDGMENT

The author(s) gratefully acknowledge the support and guidance of their supervisor and the institutional support provided by KL University, which significantly contributed to the successful completion of this research.

### REFERENCES

- [1] Bolhasani, H., Amjadi, E., Tabatabaiean, M., Jassbi, S.J. (2020). A histopathological image dataset for grading breast invasive ductal carcinomas. *Informatics in Medicine Unlocked*, 19: 100341. <https://doi.org/10.1016/j.imu.2020.100341>
- [2] Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7): 1455-1462. <https://doi.org/10.1109/TBME.2015.2496264>
- [3] Rashmi, R., Prasad, K., Udupa, C.B.K. (2022). Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review. *Journal of Medical Systems*, 46(1): 7. <https://doi.org/10.1007/s10916-021-01786-9>
- [4] Cruz-Roa, A., Basavanahally, A., González, F., Gilmore, H., et al. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *Medical Imaging 2014: Digital Pathology*, p. 904103. <https://doi.org/10.1117/12.2043872>
- [5] Yusoff, M., Haryanto, T., Suhartanto, H., Mustafa, W.A., Zain, J.M., Kusmardi, K. (2023). Accuracy analysis of deep learning methods in breast cancer classification: A structured review. *Diagnostics*, 13(4): 683. <https://doi.org/10.3390/diagnostics13040683>
- [6] Janowczyk, A., Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1): 29. <https://doi.org/10.4103/2153-3539.186902>
- [7] Srikantamurthy, M.M., Rallabandi, V.S., Dudekula, D.B., Natarajan, S., Park, J. (2023). Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. *BMC Medical Imaging*, 23(1): 19. <https://doi.org/10.1186/s12880-023-00964-0>
- [8] Madhav, B.T.P., Pardhasaradhi, P., Manepalli, R.K.N.R., Kishore, P.V.V., Pisipati, V.G.K.M. (2015). Image enhancement using virtual contrast image fusion on Fe<sub>3</sub>O<sub>4</sub> and ZnO nanodispersed decyloxy benzoic acid.

- Liquid Crystals, 42(9): 1329-1336. <https://doi.org/10.1080/02678292.2015.1050704>
- [9] Kishore, P.V.V., Venkatram, N., Sarvya, C., Reddy, L.S.S. (2014). Medical image watermarking using RSA encryption in wavelet domain. In 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, India, pp. 258-262. <https://doi.org/10.1109/CNSC.2014.6906662>
- [10] Kumar, D.A., Sastry, A.S.C.S., Kishore, P.V.V., Kumar, E.K., Kumar, M.T.K. (2018). S3DRGF: Spatial 3-D relational geometric features for 3-D sign language representation and recognition. *IEEE Signal Processing Letters*, 26(1): 169-173. <https://doi.org/10.1109/LSP.2018.2883864>
- [11] Kishore, P.V.V., Rao, G.A., Kumar, E.K., Kumar, M.T.K., Kumar, D.A. (2018). Selfie sign language recognition with convolutional neural networks. *International Journal of Intelligent Systems and Applications*, 11(10): 63. <https://doi.org/10.5815/ijisa.2018.10.07>
- [12] Kanchimani, S., Suman, M., Kishore, P.V.V. (2022). Learning global average attention pooling (GAAP) on ResNet50 backbone for person re-identification problem. *International Journal of Advanced Computer Science and Applications*, 13(7): 827-836. <https://doi.org/10.14569/IJACSA.2022.0130796>
- [13] Zhang, L., Zhang, Q., Zhao, R. (2022). Progressive dual-attention residual network for salient object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9): 5902-5915. <https://doi.org/10.1109/TCSVT.2022.3164093>
- [14] Naz, A., Khan, H., Din, I.U., Ali, A., Husain, M. (2024). An efficient optimization system for early breast cancer diagnosis based on internet of medical things and deep learning. *Engineering, Technology & Applied Science Research*, 14(4): 15957-15962. <https://doi.org/10.48084/etasr.8080>
- [15] Wang, Y.Y., Shen, H. (2026). Fine-tuning a vision-language model for automated grading of K-12 handwritten answer sheets. *Acadlore Transactions on AI and Machine Learning*, 5(2): 89-106. <https://doi.org/10.56578/ataiml050201>
- [16] Ashurov, A., Chelloug, S.A., Tselykh, A., Muthanna, M.S.A., Muthanna, A., Al-Gaashani, M.S. (2023). Improved breast cancer classification through combining transfer learning and attention mechanism. *Life*, 13(9): 1945. <https://doi.org/10.3390/life13091945>
- [17] Boumaraf, S., Liu, X., Wan, Y., Zheng, Z., et al. (2021). Conventional machine learning versus deep learning for magnification dependent histopathological breast cancer image classification: A comparative study with visual explanation. *Diagnostics*, 11(3): 528. <https://doi.org/10.3390/diagnostics11030528>
- [18] Bobowicz, M., Rygusik, M., Buler, J., Buler, R., et al. (2023). Attention-based deep learning system for classification of breast lesions—Multimodal, weakly supervised approach. *Cancers*, 15(10): 2704. <https://doi.org/10.3390/cancers15102704>
- [19] Maurya, R., Pandey, N.N., Dutta, M.K., Karnati, M. (2024). FCCS-Net: Breast cancer classification using multi-level fully convolutional-channel and spatial attention-based transfer learning approach. *Biomedical Signal Processing and Control*, 94: 106258. <https://doi.org/10.1016/j.bspc.2024.106258>
- [20] Ebrahim, M., Sedky, A.A.H., Mesbah, S. (2023). Accuracy assessment of machine learning algorithms used to predict breast cancer. *Data*, 8(2): 35. <https://doi.org/10.3390/data8020035>
- [21] Zheng, Y., Li, C., Zhou, X., Chen, H., et al. (2023). Application of transfer learning and ensemble learning in image-level classification for breast histopathology. *Intelligent Medicine*, 3(2): 115-128. <https://doi.org/10.1016/j.imed.2022.05.004>
- [22] Zou, Y., Zhang, J., Huang, S., Liu, B. (2022). Breast cancer histopathological image classification using attention high-order deep network. *International Journal of Imaging Systems and Technology*, 32(1): 266-279. <https://doi.org/10.1002/ima.22628>
- [23] Karthiga, R., Narasimhan, K., Raju, N., Amirtharajan, R. (2025). Automatic approach for breast cancer detection based on deep belief network using histopathology images. *Multimedia Tools and Applications*, 84(8): 4733-4750. <https://doi.org/10.1007/s11042-024-18949-8>
- [24] Joseph, A.A., Abdullahi, M., Junaidu, S.B., Ibrahim, H.H., Chiroma, H. (2022). Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer). *Intelligent Systems with Applications*, 14: 200066. <https://doi.org/10.1016/j.iswa.2022.200066>
- [25] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M.H., Almansour, N.A. (2021). Deep CNN model based on VGG16 for breast cancer classification. In 2021 International Conference on Information Technology (ICIT), Amman, Jordan, pp. 805-810. <https://doi.org/10.1109/ICIT52682.2021.9491631>
- [26] Yan, R., Ren, F., Wang, Z., Wang, L., et al. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, 173: 52-60. <https://doi.org/10.1016/j.jymeth.2019.06.014>
- [27] Gour, M., Jain, S., Sunil Kumar, T. (2020). Residual learning based CNN for breast cancer histopathological image classification. *International Journal of Imaging Systems and Technology*, 30(3): 621-635. <https://doi.org/10.1002/ima.22403>
- [28] Roy, S., Jain, P.K., Tadepalli, K., Reddy, B.P. (2024). Forward attention-based deep network for classification of breast histopathology image. *Multimedia Tools and Applications*, 83(40): 88039-88068. <https://doi.org/10.1007/s11042-024-18947-w>
- [29] Johny, A., Madhusoodanan, K.N. (2021). Dynamic learning rate in deep CNN model for metastasis detection and classification of histopathology images. *Computational and Mathematical Methods in Medicine*, 2021(1): 5557168. <https://doi.org/10.1155/2021/5557168>
- [30] Yu, D., Lin, J., Cao, T., Chen, Y., Li, M., Zhang, X. (2023). SECS: An effective CNN joint construction strategy for breast cancer histopathological image classification. *Journal of King Saud University-Computer and Information Sciences*, 35(2): 810-820. <https://doi.org/10.1016/j.jksuci.2023.01.017>
- [31] Toa, C.K., Elsayed, M., Sim, K.S. (2024). Deep residual learning with attention mechanism for breast cancer classification. *Soft Computing*, 28(15): 9025-9035. <https://doi.org/10.1007/s00500-023-09152-2>
- [32] Alzahrani, A. (2024). Digital image forensics: An improved DenseNet architecture for forged image detection. *Engineering, Technology & Applied Science*

- Research, 14(2): 13671-13680. <https://doi.org/10.48084/etasr.7029>
- [33] Kumar, T., Ponnusamy, R. (2023). Robust medical x-ray image classification by deep learning with multi-versus optimizer. *Engineering, Technology & Applied Science Research*, 13(4): 111406-11411. <https://doi.org/10.48084/etasr.6127>
- [34] Aswathy, M.A., Jagannath, M. (2021). An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features. *Medical & Biological Engineering & Computing*, 59(9): 1773-1783. <https://doi.org/10.1007/s11517-021-02403-0>
- [35] Singh, S., Kumar, R. (2020). Histopathological image analysis for breast cancer detection using cubic SVM. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, pp. 498-503. <https://doi.org/10.1109/SPIN48934.2020.9071218>
- [36] Thiagarajan, P., Khairnar, P., Ghosh, S. (2021). Explanation and use of uncertainty quantified by Bayesian neural network classifiers for breast histopathology images. *IEEE Transactions on Medical Imaging*, 41(4): 815-825. <https://doi.org/10.1109/TMI.2021.3123300>
- [37] Chan, R.C., To, C.K.C., Cheng, K.C.T., Yoshikazu, T., Yan, L.L.A., Tse, G.M. (2023). Artificial Intelligence in Breast Cancer Histopathology. *Histopathology*, 82(1): 198-210. <https://doi.org/10.1111/his.14820>
- [38] Elharrouss, O., Akbari, Y., Almaadeed, N., Al-Maadeed, S. (2022). Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*. <https://doi.org/10.1016/j.cosrev.2024.100645>
- [39] Kumar, G.N.K., Unhelkar, B., Vani, K.S., Chakrabarti, P., Saikumar, K. (2025). A cloud edge based heart disease detection using densenet convoluted radial basis neural network for diabetic patients. *Traitement du Signal*, 42(3): 1481. <https://doi.org/10.18280/ts.420321>
- [40] Swarnalatha, T., Supraja, B., Akula, A., Alubady, R., Saikumar, K., Prasadareddy, P. (2024). Simplified framework for diagnosis brain disease using functional connectivity. In *2024 2nd World Conference on Communication & Computing (WCONF)*, Raipur, India, pp. 1-6. <https://doi.org/10.1109/WCONF61366.2024.10692033>
- [41] Kumar, D.A., Kishore, P.V.V., Sravani, K. (2024). Deep Bharatanatyam pose recognition: A wavelet multi head progressive attention. *Pattern Analysis and Applications*, 27(2): 53. <https://doi.org/10.1007/s10044-024-01273-0>