M^{3}SCF-Net: Multi-scale Multi-Channel Multi-feature network using Resnet Based Attention mechanism for breast histopathological image classification.

Suvarna D. Pujari[1], Dr. Meenakshi M. Pawer [2], Dr. Swati P. Pawar[3,] Shital S. Pawar[4]

[1, 2,4]*Department of Electronics and Telecommunication Engineering,*
SVERIs College of Engineering, Pandharpur,
University of PAH, Solapur, India.

[3]*Department of Computer Science and Engineering,*
SVERIs College of Engineering, Pandharpur,
University of PAH, Solapur, India.

[1]sdpujari@coe.sveri.ac.in, [2]mmpawar@coe.sveri.ac.in,
[3]sppawar@coe.sveri.ac.in, [4]sspawar@coe.sveri.ac.in

Abstract -Breast cancer (BC) stands as the leading cause of mortality among women in both developing and underdeveloped nations. Precise identification of malignant lesion subtypes is critical to giving specific treatment to cancer patients. Computer-aided diagnosis (CAD) emerges as a transformative, highly accurate method for histopathological image classification. Leveraging recent advancements in computer vision and deep learning, convolutional neural networks (CNNs) have garnered remarkable success in medical image processing. Recognizing cancer subtypes automatically from whole slide images (WSI) presents a computational challenge. By training on image patches extracted from these vast images, we can design low-complexity CNNs for efficient feature extraction. Introducing the M^{3}SCF-Net (Multi-scale Multi-Channel Multi-feature network) with a ResNet-based Attention mechanism for breast histopathological image classification, we employ a knowledge-sharing strategy across the network streams and utilize attention mechanisms. Our proposed module has achieved exceptional accuracy across different magnification factors (MF) (X40, X100, X200, and X400), with peak performance observed at X200 MF, boasting an accuracy of 97.57% for multi-class and 98.95% for binary classification. We assess performance using metrics such as Accuracy, Recall, Precision, and Sensitivity, among others, to evaluate the efficacy of the proposed module in classifying histopathological breast images.Our findings reveal that the M^{3}SCF-Net model outperforms existing models such as VGG16, Xception, and ResNet152, demonstrating its superior capabilities in accurately classifying breast histopathological images.

Key Words:-Histopathological Image, Multi-scale, Multi-Channel, Multi-Featur ,DCNN

I. Introduction:

Breast cancer (BC) is one of the most often diagnosed malignancies, particularly among women globally. According to GLOBOCAN 2020, breast cancer in women outnumbers lung cancer, with 2,261,419 new cases and 684,996 deaths reported worldwide[1]. Early detection of cancer is necessary for reducing mortality rates. Breast cancer diagnosis tests include physical examination, mammography, tomosynthesis, magnetic resonance imaging (MRI), ultrasound, and biopsy [2-4]. Biopsy is regarded as the gold standard for cancer identification and subsequent treatment.The accuracy of tumour detection from histopathological images is determined by radiologists' experience; a high rate of diagnostic error is caused by oncologists' lack of experience in their field compared to experts [5-7], as well as a difference of opinion between oncologists'

explanations and expert agreement[8, 9]. Computer-aided diagnosis can help clinicians diagnose cancer at an early stage in a timely and reliable manner. The current CAD system uses hand-crafted characteristics such as wavelet coefficients, co-occurrence matrix features, and Shearlet coefficient histograms, among others, to classify malignant tissues. Many researchers are creating computer-aided tools to classify breast entire slide images as non-cancerous (benign) or cancerous (malignant) [10]. Haralick, wavelet-based, intensity-based, and morphological characteristics are employed to extract relevant features from segmented nuclei and their environs in order to classify histopathology images [11], [12]. Due to the complicated structure of histopathology pictures, the handmade feature extraction technique is ineffective [13].Breast cancer is not a single illness, but rather a collection of subgroups with distinct molecular aetiology and clinical characteristics [14], [15]. As a result, recognising cancer subtypes is extremely challenging in order to facilitate accurate cancer diagnosis and therapy. Deep neural network creation has recently become easier due to the availability of powerful computational tools and vast training databases [16]. Deep learning (DL) uses nonlinear transformations to extract high-level abstractions directly from images [17]. Breast cancer detection relied on supervised machine learning methods, including SVM, K-NN, and Naïve Bayesian [18], [19].Breast cancer is not a single illness, but rather a collection of subgroups with distinct molecular aetiology and clinical characteristics [14], [15]. As a result, recognising cancer subtypes is extremely challenging in order to facilitate accurate cancer diagnosis and therapy. Deep neural network creation has recently become easier due to the availability of powerful computational tools and vast training databases [16]. Deep learning (DL) uses nonlinear transformations to extract high-level abstractions directly from images [17]. Breast cancer detection relied on supervised machine learning methods, including SVM, K-NN, and Naïve Bayesian [18], [19].Our main contribution is to improve the accuracy of classification of histopathological patches created [25] from whole slide images (WSI), different filter size gives the multiple feature maps and these feature through CBMA integrated with skip connections gives Robust and Refined features.

II. Related Studies:

In another study by Zanariah et al. [26], Three Deep Layer Convolution neural network (DCNN) Architectures were used to detect breast cancer. In this method, they initially created patches from whole slide image which Mitosis and non- Mitosis were selected of size 64 x 64 x 3.These patches were given to 6-layer CNN, 13-layer CNN 17-layer CNN and 19-layer CNN, these three different CNN architecture were was performed for CNN-based algorithms for 5-fold. They achieved an accuracy of 84.49% by using 19-CNN layer, which was the highest among three CNN.

Wang et al. [21] proposed a deep learning-based bilinear CNN method to classify histopathological images. The convolution and pooling layers applied on hematoxylin and eosin(h & e) stained breast tissue images decompose into h & e parts and then on both of these images combined and learn more effective feature representation. These two features were fused by bilinear pooling. This proposed method got success of an average accuracy of 92.6±1.2% across all folds to classify the H&E stained breast cancer images.

Gandomkar et al.[20], the major findings of this research work: (i) To classify histopathological breast images as cancerous (malignant) or non-cancerous (benign), (ii) To perform sub-classification of cancerous (malignant) images as DC, LC, MC, PC; and (iii) To perform sub-classification of non-cancerous (benign) images as A, F, PT, or TA. In this study, Multi-category classification of breast histopathological image using Deep Residual Networks (MuDeRN) was a two-stage network approach used to classify patients with the use of H & E stained slides. At first stage histopathological breast image is detected either benign or malignant, and then at second stage malignant and benign cancer were categorizing for each class into four subtypes. MuDeRN was a very Deep CNN comprises 152 layers (ResNet152). The proposed framework was trained for different MF× 40,× 100,× 200 and × 400. First stage of MuDeRN achieved classification rates (CCR) of 98.52%, 97.90%, 98.33% and 97.66% in × 40,×

100,× 200 and × 400 MF, respectively. For the eight-class categorization of images based on the output of MuDeRN in second stage, Correct Classification Rates were 95.40%, 94.90%, 95.70% and 94.60% in × 40,× 100,× 200 and × 400 respectively.

The research study conducted by SanaUllah et al. [27], developed deep learning approach using CNN called BreastNet. BreastNet module comprised with convolutional block attention module (CBAM), Skip connections, dense block, and hyper-column technique. This model was a multi-classification of breast histopathological image. They also trained the proposed model for distinct magnification factors (× 40,× 100,× 200 and × 400).

Kalpana George et al. [28] for automated breast cancer detection proposed method extracts feature from a nucleus using a DCNN i.e. 'NucDeep',. They designed simple CNN with less complexity for feature extraction from non-overlapping nuclei patches spotted over the images. A feature fusion (FF) approach with SVM classification framework was applied to categorize breast histopathological images from the extracted Convolution Neural Network features from nuclei. The FF method converts the local features from nuclei into a compact feature at image-level that improves the classifier performance. The success rate for cancer detection was recognition rate of 96.66 ± 0.77 % , specificity of 98.50% and sensitivity 96.21%.

III. Contribution:

In various medical image applications, CNN models have achieved promising results [20, 22]. CNN is gives the more robust and refined features. The aforementioned studies indeed motivated us to explore the performance of different deep CNNs and also to validate the accuracy and network parameters for assessment of breast cancer using histopathological images.

The major contributions of this study are given below:

1. The proposed a multi-scale filter bank with an attention mechanism for feature extraction.
2. The proposed network consists of the CBMA block, to examine the feature maps: channel and spatial and to get refined features.
3. The CBMA is integrated with ResNet(Skip Connections) Block in Resnet.
4. The proposed network consists of multi-streams for fine-level as well as course-level feature extraction.
5. The proposed network follows a knowledge sharing strategy by sharing learned features at each stream across the network.
6. Our developed model is light weight model.
7. The proposed network is trained and validated for histopathological hematoxylin-eosin stained breast digital slides on BreakHis dataset.

The remaining the paper is organized as follows. Section IV gives details' of the proposed approach and the materials used. Section V offers the experimental results and discussions of the proposed network architecture. Finally, in Section VI refers concluding remark.

IV. MATERIAL AND METHODS:

This section discusses the proposed approach for classification of breast histopathological images; further it provides a brief explanation of the dataset used for our work. As with the traditional pattern recognition pipeline, we extract robust and refined features followed by pattern classification. As we know, effective feature extraction is a key to get higher accuracy. As discussed in previous sections, the existing handcrafted feature extractor fails to extract robust features when there is a complex image structure. In the last decade, the convolution neural network's feature learning ability attracted many researchers to incorporating CNN for feature extraction. Thus, the CNN's robustness in feature extraction inspired us to propose a CNN network for Breast image classification.

A. BREAKHIS Database

BREAKHIS DATABASE contains 7,909 microscopic breast cancer histopathological RGB images of size 700 X 460 pixels from 82 patients with four visual MLs - × 40,× 100,× 200 and × 400 as shown in Fig 1. Here, benign class consists of 2,480 images and malignant class of 5,426 images. The BREAKHIS database built-in P&D Laboratory

Pathological Anatomy and Cytopathology, Parana, Brazil [29, 30].

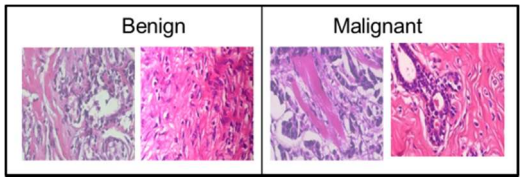

Fig. 1. Sample images from BreakHis Dataset.

As shown in Table 1, Benign are represented using four different subtypes as Adenosis(A), Fibroadenoma(F), Tubular Adenoma (TA), Phyllodes Tumor(PT). Whereas, Malignant subclasses are Ductal Carcinoma(DC), Lobular Carcinoma(LC), Mucinous Carcinoma(MC), Papillary Carcinoma (PC). The details about BreakHis database is given in Table 1.

Table 1.The BreakHis Database distributed into four magnification levels and main classes, their subcategories

Class	Sub-Classes	Magnification Level 40X	100X	200X	400X	Total
Benign	A	114	113	111	106	**444**
	F	253	260	264	237	**1014**
	TA	109	121	108	115	**453**
	PT	149	140	150	130	**569**
Malignant	DC	864	903	896	788	**3451**
	LC	156	170	163	137	**626**
	MC	205	222	196	169	**792**
	PC	145	142	135	138	**560**
Total		**1995**	**2081**	**2013**	**1820**	**7909**

From Table 1, it can be noted that at magnification level$\times$ 40, the database contains total 1995 histopathological images. Next, at $\times$ 100 it has 2081 images, at $\times$ 200 it consists of 2013 images and finally at $\times$ 400it represents 1820 images.

B. Methodology

In this work, we proposed Multi-scale, Multi-Channel Multi-feature network using Resnet Based Attention mechanism for breast histopathological image classification for training and prediction of breast cancer as shown in Fig2.The proposed network made up of three streams belongs to a trail of convolution layer having filter size of $1 \times 1, 3 \times 3$ and 5×5 respectively on CBMA [31] layer. Each Convolution layer processes the patch from of input breast histopathological image and extracts the features at a particular scale; before the convolutional block, we use the CBMA block, which gives the focused region. For better feature learning, share the learned feature maps within the network stream. This inspired by the work [32].

Element-wise addition and concatenation are two ways to share/integrate the learned feature maps. The feature map concatenation increases the number of network trainable parameters. Thus, keeping in mind the real-time use case of the proposed network, we use the element-wise addition operation to share the learned feature maps within the network streams. Fig 2 shows, the feature map concatenation spikes the number of network trainable parameters. Thus, keeping in mind, we make use of element-wise addition operation to share the learned.

C. Preprocessing:

a. Patch Creation:

The whole slide image (WSI) of Giga-pixel resolution is computation hard to GPUs and cannot fit in GPU memory. There are three patch selection approaches i) Random selection, it is simplest approach of patches from WSIs images are randomly [25]. ii) To Identify cancerous lesion in the image, machine learning algorithm for detection of cancer can be trained by randomly selected patches which can be patch annotated lesion for cancerous part in WSIs by radiologist [33] or non-annotated patch as non-cancerous WSIs. iii) Cluster based image patches approach, clustering of image patch is used to obtain variegated appearance of breast tissues. The cluster operation is carried out over 8 subtype image patches to represent WSI [34]. To identify subtype from image patch across whole training dataset centroid of image patch is calculated. Further, centroid of the 8 subtype clusters from database is compared and closest centroid of the cluster will be assign to the image patch. Along the side this process is repeated iteratively for every epoch of the training [35]. In this work due to simplicity we created patches using Random selection approach.

b. Augmentation:

Generally, augmentation was done on training data for better performance of network and keeps the network away from overfitting. Each image is rotated with different angles like 1800, 2700 or 900 and flipped with respect to horizontal or vertical axes along with random shift about ± 10 pixels. While training of network for every epoch patches from images were augmented randomly and provided to the network.

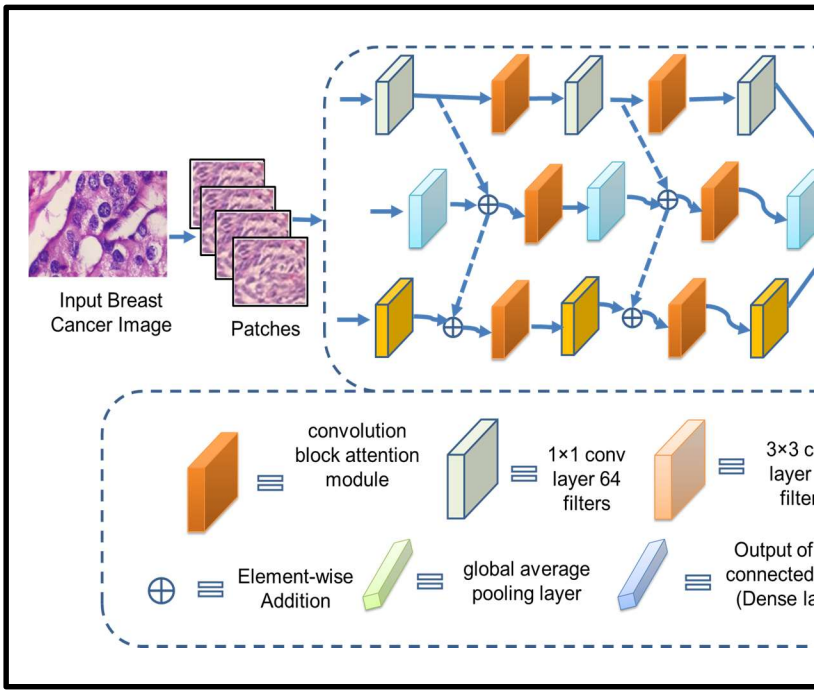

Fig.2. Block diagram of the proposed MuSCF-Net with Resnet based attention mechanism.

1) Convolution of filter bank for feature extraction:

We used the CNN's basic building block, such as convolution, Relu and batch normalization, for feature extraction with different filter size. Fig 2 shows the block diagram of the proposed multi-scale multi-stream with attention mechanism deep neural network. It is very difficult to choose the filter size for a particular convolution layer, and there is no universal rule for filter size selection. Thus, we choose filter size (1×1, 3×3, 5×5), multi-scale filter bank for feature extraction. This multi-scale filter bank helps tackle network design difficulties by letting the network decide the best route for itself.

2) Convolution Block Attention Module (CBMA) integrated with ResBlock in *CBMA Integrated with ResNet Blockin Resnet:*

As illustrated in Fig 3, the CBMA divided into sub models-channel sub-module uses both max-pooling outputs and average-pooling outputs with a shared network; this shared network is made up of multilayer perceptron (MLP) with one hidden layer. The Skip connections gives the class wise input to the output.

Channel Attention (CA) module

The CA module produces feature map for each Channel. It stresses on important parts of an input image. In CA module, average and maximum pooled

features are extracted and processed together [31], [27]. In this module, average and max pooling operations are used to combine spatial information of a feature map, which produces two spatial context descriptors: F^c_{max} and F^c_{avg}, as max-pooled features and average -pooled features respectively. Then these spatial context descriptors are provided to the given shared network. This will generate the CA map as $M_c \in \mathbb{R}^{C\times1\times1}$. In this shared network, it consists of one hidden layer of MLP. The parameter overhead is reduced by setting hidden activation size as $\mathbb{R}^{C/r\times1\times1}$, here r denotes reduction ratio. The element-wise summation operation is performed on output feature vectors with the application of shared network over each spatial context descriptors.
Accordingly the CA can be calculated as follows:

$$M_c(F) = \sigma\,(\mathrm{MLP}\,(\mathrm{AvgPool}\,(F)) + \mathrm{MLP}\,(\mathrm{MaxPool}\,(F))) \quad (1)$$

$$= \sigma\,(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \quad (2)$$

Where σ the sigmoid function, $W_0 \in \mathbb{R}^{C/r\times C}$, and $W_1 \in \mathbb{R}^{C\times C/r}$. Note that the multilayer perceptron weights, W_0 and W_1, are shared for both inputs. The activation function (ReLU) is followed by W_0.

Spatial Attention (SA) module

Along the channel axis in spatial sub-module two similar outputs pooled, concatenating them to generate an efficient feature descriptor and forwarding them to a convolution layer. The CBMA module not only gives the refine feature map but robust to the noise input. It decomposes the process into a channel and spatial attention separately due to separation reduction in computation parameters [31], [27].

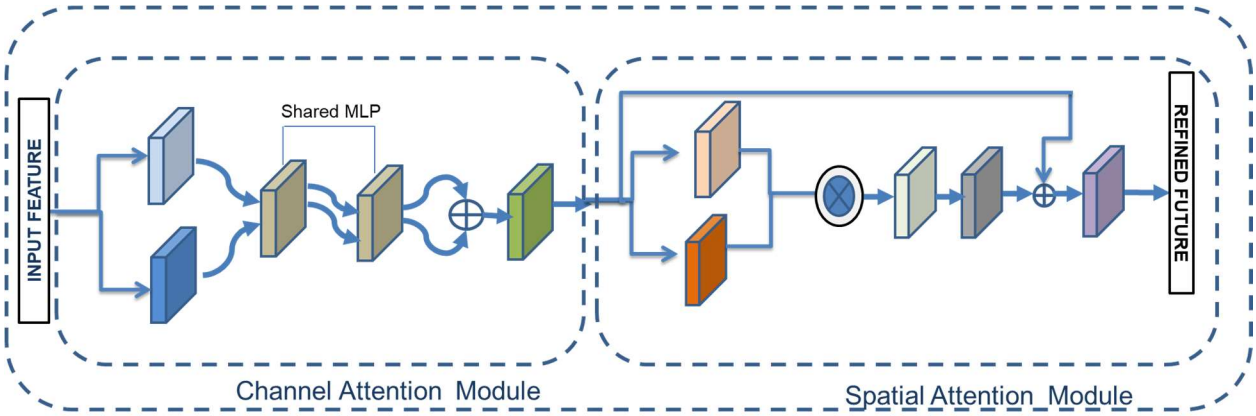

Fig.3. Convolution Block Attention Model (CBMA) integrated with Skip connections in Resnet

The highlighting informative regions are generated across channel axis with the use of pool operation over concatenated feature frames. The spatial attention map $M_s(F) \in R^{H\times W}$ is generated by applying a convolution layer which encodes where to stress or suppress. Details of the operations have shown below eq (3) and (4). By using two pooling operation we aggregated channel information of a feature map and generated two 2D maps: $F^c_{avg} \in \mathbb{R}^{1\times H\times W}$, and $F^c_{max} \in \mathbb{R}^{1\times H\times W}$ across the channel features as 1. Average pooled feature map and 2. Max pooled feature map. By a standard convolution layer these are then concatenated and convolved to produce 2 dimensional (2D) SA map. Accordingly the SA can be calculated as follows:

$$M_s(F) = \sigma\,(f^{7\times7}([\mathrm{AvgPool}(F);\ \mathrm{MaxPool}(F)]) \quad (3)$$

$$= \sigma\,(f^{7\times7}([F^c_{avg};\ F^c_{max}]) \quad (4)$$

Where σ represents the sigmoid function, $f^{7\times7}$ represents a convolution operation with the filter size of 7 × 7. These features are then transferred to CNN. Two attention modules, CA and SA compute complementary attention, focusing on 'what' and 'where' respectively. These two modules placed sequentially.
The training of DNN becomes difficult due to vanishing gradient problem. To tackle this problem, the researchers have increased depth of the network

by staking layers together. It is observed that this technique is not suitable and as network goes deeper and deeper the performance of network decreases rapidly sometimes network saturates.
Therefore identity mapping with the use of short connections have been proposed [24], in this work the author has introduced skip connections which skips layers and joins directly, where the connection is provided this network is known as Resnet. The idea of skip connection while training, where it skips some layers and joins directly output layer. The major benefit of using skip connection is performance degrading layer in the architecture can be skipped by regularization.

The vanishing gradient problem is overcome by introducing skip connection i.e. deep residual learning framework in DNN. Let's H(x) can be a non-linear function which represents the mapping from input to output, F(x):= H(x) –x instead of F(x) +x.

Global Average Pooling Layer

The global average pooling layer used instead of the traditional, fully connected layer in CNN [32]. The last multi-perceptron convolution (mlpconv) layer generates one feature map for each corresponding class of the classification task; this feature map is a confidence map of classes due to mlpconv. The resulting vector is the average of each feature map and forwarded to the softmax layer shown in Fig 2. Due to no parameter to optimize over fitting is avoided in the global average pooling; thus, in this layer.
Let, x; F represents the convolution layer's input and output in the proposed network, respectively.

$$\text{Thus, } F_{ijN} = \sum_{uvC} W_{uv}CN \times X_{i+u,j+v,C} \quad (5)$$

Where, $W_{uv}CN$ denotes the filterbank having filters of spatial size $M \times N$ and locations

$$u \in \left[-\frac{(M-1)}{2}, \frac{(M-1)}{2}\right], \quad v \in \left[-\frac{(K-1)}{2}, \frac{(K-1)}{2}\right]$$

C, N the number of channels and filters respectively (i; j) denotes the image location. After instance normalization go through the ReLu as follows

$$F_{1ijk} = \frac{F_{ij}-\mu_k}{\sqrt{\sigma^2+\varepsilon}} \quad (6)$$

$$\mu_k = \frac{1}{HW}\sum_{i=1}^{W}\sum_{j=1}^{H} F_{ijk} \ , \sigma_k^2 = \frac{1}{HW}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(F_{ijk}-\mu_k\right)^2 \quad (7)$$

Where, $\mu_k$and $\sigma_k$represent the mean and variance of the k^{th}feature map, W and H denotes the width and height of the feature map.

$$F_{2ijN} = \max\{0, F_{1ijN}\} \quad (8)$$

Dropout Layer

The dropout layer is used for regularization in CNN and avoids the overfitting issue. It temporarily removes some units from the network and all its incoming and outgoing connections[36], used in the M^{3}SCF-Net model of 0.2 dropout.

Dense Block

Dense layer is widely used in the neural network and used to change the vector dimension. In this layer, every neuron receives input from all neurons of its previous layer. So it is called a deeply connected layer. In the MuSCF-Net model three Dense Block i.e. fully connected blocks used.

D. Training Details

M^{3}SCF-Net Network trained on BreakHis dataset [23] Details of the dataset given in Table1. We created 64 x 64 size patches of an image from a given dataset and considered 80% patches as training and 20% as a validation set of each class; each magnification level. We trained our proposed model for x40, x100, x200, and x400 magnification factor.
Used an Adam Optimizer and loss function as categorical cross-entropy to train a proposed breast histopathological image classification network. The network is trained on *NVIDIA GeForce RTX 2060 super* GPU for 10 fold cross-validation in each fold 10 epochs. With a learning rate – *Min* 0.0001 to *Max* 0.001. We used Keras library on Tensorflow to design and train the proposed network.

V. RESULTS AND DISCUSSION

In this section, we have discussed the experimental analysis of the M^{3}SCF-Net for breast histopathological image classification. For a classification problem, accuracy is not efficient as a performance measure; an alternative for classification accuracy is to use precision and recall metric. The performances measure parameters of the network like accuracy, precision(Pr), sensitivity(Se),

specificity(Sp) and f1-score is shown in equation (11),(12),(13) and (14), respectively[39]. The formulations of the matrices are described as follows:

Table 3: The accuracy and loss of existing CNN models and the proposed M^{3}SCF-Net.

Model		Acc(%)	Loss (%)	Computational Parameter in million
VGG16	Multi-class	81.37	20.48	138
	Binary	92.58	10.20	
Xception	Multi-class	96.57	8.6	22.9
	Binary	99.01	1.68	
Proposed Model	Multi-class	**97.57**	**3.01**	**0.58**
	Binary	**98.92**	**1.09**	

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Se/Recall} = \frac{TP}{TP+FN} \quad (13)$$

$$\text{F1 Measure} = \frac{(2\times Precision\times Recall)}{(Precision+Recal\ )} \quad (14)$$

Here,

TP (true positive) gives the number of malignant and their subclass images accurately predicted as malignant,

TN (true negative) gives the number of benign and their subclass images accurately predicted as benign,

FP (false positive) gives the number of benign and their subclass images inaccurately predicted as malignant,

FN (false negative) gives the number of malignant and their subclass images inaccurately predicted as benign.

The first part, the experimental result of our proposed method were compared with existing models like VGG16 and Xception model shown in Table 3, and the MuSCF-Net model achieved a superior i.e. 98.07% and also less computational parameter 0.58 million result than the existing CNN models. We got results at different magnification levels (X40, X100, X200, and X400) of images but the proposed network gave best result at x200. The BreakHis contains benign and malignant and each into four subclasses. Table 4 examined the each subclasses performance measure; the average classification accuracy of subclasses of benign data is 98.13%. Similarly, the average classification accuracy of malignant data is 97.01%.

These superior results achieved by proposed MuSCF-Net model 8due to robust feature from Convolution of filter bank and then refined in CBMA. Multi-scale feature extraction as well as multi-stream network architecture and feature sharing strategy gives the robust features, refines them channel attention and spatial attentions models.

Table 4. Performances measure parameters of proposed M^{3}SCF-Net module for benign and malignant images at X200 Magnification level.

Class	Sub-Classes	Acc (%)	Pr (%)	Recall (%)	F1 SC (%)	Avg Acc (%)
Benign	A	98.54	98.20	99.20	98.70	97.75
	F	97.07	97.10	98.14	97.62	
	PT	97.59	96.01	92.02	93.97	
	TA	99.01	99.33	99.63	99.47	
Malignant	DC	95.30	95.40	97.43	96.32	96.51
	LC	96.03	86.10	79.87	82.95	
	MC	99.15	97.01	96.55	97.32	
	PC	97.37	97.50	97.54	97.88	

VI. CONCLUSION

In this work, we proposed the DCNN for H & E stained breast histopathological image classification. The proposed Multi-scale, Multi-Channel feature network using Resnet Based Attention mechanism for breast histopathological image classification consists multi-scale filter bank and composed of multiple streams for fine and cross level feature with attention mechanism to extract the robust features, and features are shared in between streams and ResBlock from Resnet. This study focused on improving classification accuracy and the classification has been carried out on BREAKHIS dataset. Patches of size 64 x 64 x 3 are created from histopathological images of BREAKHIS dataset. The classification accuracies and performance measure parameters of proposed model have superior at 200Xmagnification level than previously available models. This model is light weight model of lower computational parameter 0.58 million.

In BREAKHIS dataset only benign and invasive (malignant) cancer types were included other cancer types such as HER_2, in situ cases etc. we're not included. We use the different dataset to train the MuSCF-Net model. The proposed module can be generalized to the design of high-performance CAD systems for other bio-medical imaging processing

tasks, in the future.

REFERENCES

1. Sung, H., et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 2021.
2. Nover, A.B., et al., Modern breast cancer detection: a technological review. International Journal of Biomedical Imaging, 2009. **2009**.
3. Yang, Z., et al., EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. Neurocomputing, 2019. **366**: p. 46-53.
4. Screening, P. and P.E. Board, Breast Cancer Screening (PDQ®), in PDQ Cancer Information Summaries [Internet]. 2020, National Cancer Institute (US).
5. Allison, K.H., et al., Trends in breast biopsy pathology diagnoses among women undergoing mammography in the United States: a report from the Breast Cancer Surveillance Consortium. Cancer, 2015. **121**(9): p. 1369-1378.
6. Hamidinekoo, A., et al., Deep learning in mammography and breast histology, an overview and future trends. Medical image analysis, 2018. **47**: p. 45-67.
7. Allison, K.H., et al., Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. Histopathology, 2014. **65**(2): p. 240-251.
8. Elmore, J.G., et al., Variability in pathologists' interpretations of individual breast biopsy slides: a population perspective. Annals of internal medicine, 2016. **164**(10): p. 649-655.
9. Coracin, F., et al., DIAGNOSTIC CONCORDANCE AMONG PATHOLOGISTS INTERPRETING ORAL MUCOSAL BIOPSIES FROM INDIVIDUALS AFFECTED BY GVHD. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 2019. **128**(1): p. e36-e37.
10. Gandomkar, Z., P.C. Brennan, and C. Mello-Thoms, Computer-based image analysis in breast pathology. Journal of pathology informatics, 2016. **7**.
11. Niwas, S.I., P. Palanisamy, and K. Sujathan. Wavelet based feature extraction method for breast cancer cytology images. in 2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA). 2010. IEEE.
12. Weyn, B., et al., Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. Cytometry: The Journal of the International Society for Analytical Cytology, 1998. **33**(1): p. 32-40.
13. Liu, Z., X.-S. Zhang, and S. Zhang, Breast tumor subgroups reveal diverse clinical prognostic power. Scientific reports, 2014. **4**(1): p. 1-9.
14. Guo, Y., et al., Improvement of cancer subtype prediction by incorporating transcriptome expression data and heterogeneous biological networks. BMC medical genomics, 2018. **11**(6): p. 87-98.
15. Guray, M. and A.A. Sahin, Benign breast diseases: classification, diagnosis, and management. The oncologist, 2006. **11**(5): p. 435-449.
16. Bengio, Y., Learning deep architectures for AI. 2009: Now Publishers Inc.
17. Ker, J., et al., Deep learning applications in medical image analysis. Ieee Access, 2017. **6**: p. 9375-9389.
18. Gupta, P. and S. Garg, Breast cancer prediction using varying parameters of machine learning models. Procedia Computer Science, 2020. **171**: p. 593-601.
19. Akay, M.F., Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 2009. **36**(2): p. 3240-3247.
20. Gandomkar, Z., P.C. Brennan, and C. Mello-Thoms, MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. Artificial intelligence in medicine, 2018. **88**: p. 14-24.
21. Wang, C., et al. Histopathological image classification with bilinear convolutional neural networks. in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017. IEEE.
22. Hasan, S. Deep Layer CNN Architecture for Breast Cancer Histopathology Image Detection. in The International Conference on Advanced Machine Learning Technologies and

Applications (AMLTA2019). 2019. Springer.
23. Pujari, S.D., M.M. Pawar, and M. Wadekar, Multi-Classification of Breast Histopathological Image Using Xception: Deep Learning with Depthwise Separable Convolutions Model, in Techno-Societal 2020. 2021, Springer. p. 539-546.
24. He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
25. Hou, L., et al. Patch-based convolutional neural network for whole slide tissue image classification. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
26. Zainudin, Z., S.M. Shamsuddin, and S. Hasan. Deep layer CNN architecture for breast cancer histopathology image detection. in International Conference on Advanced Machine Learning Technologies and Applications. 2019. Springer.
27. Toğaçar, M., et al., BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. Physica A: Statistical Mechanics and its Applications, 2020. **545**: p. 123592.
28. George, K. and P. Sankaran, Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features. Computer Methods and Programs in Biomedicine, 2020. **194**: p. 105531.
29. Benhammou, Y., et al., BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. Neurocomputing, 2020. **375**: p. 9-24.
30. Szegedy, C., et al. Going deeper with convolutions. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
31. Woo, S., et al. Cbam: Convolutional block attention module. in Proceedings of the European conference on computer vision (ECCV). 2018.
32. Lin, M., Q. Chen, and S. Yan, Network in network. arXiv preprint arXiv:1312.4400, 2013.
33. Feng, Z., J. Yang, and L. Yao. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. in 2017 IEEE International Conference on Image Processing (ICIP). 2017. IEEE.
34. Yao, J., et al., Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Medical Image Analysis, 2020. **65**: p. 101789.
35. Xie, C., et al. Beyond classification: Whole slide tissue histopathology analysis by end-to-end part learning. in Medical Imaging with Deep Learning. 2020. PMLR.
36. Srivastava, N., et al., Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 2014. **15**(1): p. 1929-1958.
37. Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
38. Ho, Y. and S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE Access, 2019. **8**: p. 4806-4813.
39. Powers, D.M., Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061, 2020.