RESEARCH ARTICLE

MSF-Model: Multi-Scale Feature Fusion-Based Domain Adaptive Model for Breast Cancer Classification of Histopathology Images

HAMEED ULLAH KHAN[1], BASIT RAZA[1], ABDUL WAHEED[2,3], AND HABIB SHAH[4]
[1]Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
[2]Department of Computer Science, Women University Swabi, Swabi 23430, Pakistan
[3]School of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea
[4]Department and College of Computer Science, King Khalid University, Abha 62529, Saudi Arabia

Corresponding author: Basit Raza (basit.raza@comsats.edu.pk)

ABSTRACT One of the most common causes of mortality for women globally is breast cancer. Early breast cancer identification could make it possible for people to receive the appropriate treatment to save their lives and return to their routine lives. Breast cancer diagnosis by histopathology is referred to as the gold standard. In recent years, convolutional neural network-based techniques are used for breast cancer classification. However, they faced domain adaptation, small objects retention, and feature extraction issues of complex microscopic images. In this study, we introduced multi-scale feature fusion-based domain adaptive model for breast cancer classification using histopathology images. It has two blocks and six lightweight sub-models where each block contains three models. Dilated layers are used in sub-models to overcome the disappearing of small objects in deep layers. Reducing the disappearing of small objects helped to extract better features for higher performance. Multiple heterogeneous feature extractors are used in this study which helped to extract various features. Extracted features are fused and reduced by retaining better features. Learning of model from natural images to complex microscopic images has limitation of domain adaptation. Same domain transfer learning is used in this study to overcome the limitations of different domain transfer learning. Model is trained on patchcamelyon17 dataset and weights of this training are further used for same domain transfer learning. Pre-trained weights are further used for the training of proposed model on BreaKHis dataset. A number of conventional data augmentation techniques are used as complex models require higher number of samples for the tuning of weights. Local window based CLAHE contrast enhancement technique is used to increase foreground-background contrast and remove noise. The proposed model achieved 98.00% precision, 98.15% recall, 98.08% f-measure, and 98.23% accuracy on test data. To best of our knowledge, it surpassed state-of-the-art models.

INDEX TERMS Deep learning, convolutional neural networks, breast cancer classification, histopathology images.

I. INTRODUCTION

According to GLOBOCAN 2020, female breast cancer is the most common cancer type which surpassed the previous most common lung cancer. It is 11.7% of all types of cancer cases in women with 2.3 million new cases. In terms of mortality it is the fifth death causing disease in the world with 685,000 number of deaths [1].

Some well-known techniques: fine needle aspiration cytology (FNAC) [2], magnetic resonance imaging (MRI), mammography and histopathology are used for screening and diagnosis of cancer. A pre-trained CNN models were

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen.

used [3], [4] for cancer analysis on mammogram images. Histopathological images are most widely used for classification [5]. Histopathology is considered as a challenging task [6] due to its complexity like cell overlapping, unusual color distribution and high similarity in images [7]. In histopathology, tissues of patient are taken and then hematoxylin and eosin (H&E) stain is applied to highlight nuclei and other structure of tissues, respectively. It is implemented to examine tissues of histopathology images at various magnification levels taken through microscopes.

Pathologists apply various manual, time consuming and costly techniques to correctly diagnose the samples of cancerous regions [8]. To overcome these problems, whole slide histopathology images having information of whole image or photo-microscope images of small portion of a slide are taken [9] and computer-aided diagnosis (CAD) systems are applied. Deep learning models like convolutional neural networks (CNNs) which are most rapidly used deep learning models for image feature extraction and classification [10] are used. These models motivated the researchers to develop fast and accurate computer-aided diagnosis (CAD) systems for different machine learning and image processing applications. These models are mostly used in age assessment [11], diabetic retinopathy screening [12], [13], lung cancer [14], breast cancer [15], skin cancer [16], brain tumor [17], cervical cancer [18], liver cancer [19], bone disease prediction [20].

Deep learning techniques are used to overcome manual qualitative analysis issue. However, several deep learning models faced some challenges during classification of breast cancer histopathology images. One of them is that histology images have high resolution, so these are divided into patches which reduces accuracy of classifier. A classifier may be good for one or more patches, but it is not necessary for it to be best for whole slide histopathology image. Another issue of histopathology images is that features of one patch do not fully represent the whole image, so image-wise fusion loses large amount of information. Dataset is also a challenge for histopathology images. Medical images datasets are not as publicly available as these are required for deep learning models. Some datasets which are publicly available have small amount of data [5]. So, main challenge for researchers is to access database having efficient and large data to increase validation accuracy [8]. Researchers apply different machine learning and deep learning techniques for data augmentation. In histopathology it is tedious and time consuming to identify images due to their complexity. Computational techniques have high false positive and false negative [21] due to low foreground-background contrast and noise during staining process. Some other factors are touching boundaries [22] and mitotic and non-mitotic cells similarities [23] are the reasons of low accuracy. Another factor of low accuracy is disappearing of small objects in deep layers of histopathology images due to image resizing. Models are unable to extract features of disappeared objects which degrades their performance.

In this study, our objective is to extract features of images based on multi-scale inputs. We used heterogeneous models to extract different features for better classification of complex images. Mostly, models fail due to less amount of data. Our objective in this study, is to train models on same domain to get better weights. Computational cost of deep learning models is observed as high however, we used lightweight deep learning models. The contributions of the study are as follows:

- Several studies identified that performance of multiple models remains high as compared to stand-alone models in complex microscopic images. Multi-scale feature fusion of lightweight heterogeneous models is introduced in the proposed MSF model to extract cell and tissue level features. Strengths of six independent lightweight models are fused in term of features to classify breast cancer histopathology images.
- Dataset is one of the key factors of the performance of deep learning model. Data is augmented twelve times to the original data by using conventional data augmentation techniques. CLAHE is used for foreground-background contrast enhancement of complex histopathology images.
- Transfer learning techniques are used for fast and better performance of deep learning models in image processing. It is observed from the literature that transfer learning from natural images to complex histopathology images faces domain adaptation challenge. In this study, same domain transfer learning is used to retrain the model by using pre-trained weights of patchcamelyon17 dataset.
- Performance of deep learning models degrade in complex histopathology images due to disappearance of small microscopic objects in deep layers. It occurs due to image dimensionality reduction in pooling layers and convolutional layers having stride more than one. Dilated layers in the six models are used to reduce the disappearing chance of small objects in deep layers. The proposed model outperformed the existing state-of-the-art models.

Rest of the paper is organized in such a way that section II is about related work, dataset is described in section III. The proposed multi-scale feature fusion model is explained in section IV. Section V is discussing results while ablation study is discussed in section VI. Conclusion with future work is enlightened in section VII.

II. RELATED WORK

VGG-16 model [24] is used for deep feature extraction. These features are fed to Support Vector Machine and Random Forest as input. Different augmentation techniques like rotation, mirror projection as left-right and top-bottom, gaussian blur and positive scaling are used. Overall, these augmentations techniques increased datasets up to 12 times to the original datasets. In this study, 5-fold technique is applied for

training and testing phase and model achieved 97% accuracy, 98.9% area under the curve (AUC). In another study [25], Convolutional neural network is used to take RGB images of dimension 32 × 32 or 64 × 64 pixels as input for binary classification. Images are divided into patches using sliding window with 50% overlapping and random patches extraction with no overlapping. To improve accuracy, sum, product and max fusion rules are applied. Finally, model achieved 90% patient level, and 85.6% image level accuracy for 40× magnified images of BreaKHis dataset using max fusion rule.

ResHist model [7], inspired by deep residual learning is proposed for breast cancer histopathological images classification. It has 152 layers including 13 residual blocks. This is fully automatic model with no pre-processing. BreaKHis dataset is used for validation of the model. To avoid overfitting, dataset is augmented using various data augmentation techniques like stain normalization, patches generation and affine transformation. These augmentation techniques increased the dataset eleven times to its standard size. Model achieved f1-score to 93.45% and accuracy up to 92.52%. In study [26], combination of transfer learning, deep learning and GAN is introduced which gained accuracy up to 98.1%.

Several transfer learning techniques [9], [27], [28] are used to get better results on smaller number of epochs. In study [9], ten pre-trained CNNs models are utilized for feature extraction of histopathology images. These different pre-trained models are fed with input images of 224 × 224 × 3 to 299 × 299 × 3 pixels resolution where actual size of 700 × 460 × 3. Six non-overlapping patches of size 224 × 224 × 3 are given as input to pre-trained models. Pre-trained VGG16, VGG19, AlexNe, Inception-v3, GoogLeNet, inception ResNetV2, ResNet18, ResNet50, and ResNet101 and SqueezeNet are used. These models are used for feature extraction. After feature extraction through these models, six different classifiers which are Linear SVM, fine KNN, cosine KNN, fine tree, bagged tree, boosted tree are used. Model achieved up to 89% patient recognition rate. Similarly in [28], same and different domain datasets are used to pre-train hybrid convolutional neural network model. As compared to different domain transfer learning, in same domain transfer learning model achieved higher accuracy as model learned better features from similar types of images.

Parallel combination of DenseNet and a recurrent neural network (RNN) is introduced [29] for RGB features extraction. In this study, switchable normalization (SN) technique is used for image normalization which combined batch normalization, layer normalization and instance normalization. Proposed model gained an accuracy of 92%, 98.3% and 97.5% on BATCH2018, Bioimaging2015 and extended Bioimaging2015 datasets, respectively. Similarly, convolutional neural network (CNN) classification model is proposed [8] which has two branches; single-task classifier and multi-task classifier. Single-task classifier is just used to classify input images into cancerous and non-cancerous. On the other hand multi-task classifier provides magnification level information along with classification. Model has average accuracy

of 83.25% which is not as good as to use for clinical purpose. FE-BkCapsNet [30] is introduced, which is the combination of CNN and capsule network (CapsNet). Purpose of CNN is to highlight semantic and purpose of capsule network is to give information about position of objects.

In study [31], GoogLeNet, VGGNet and ResNet models are used for feature extraction. Extracted features are provided to a fully connected layer for binary classification of histopathology images using average pooling. The proposed model gained an accuracy of 97.525% for two classes: benign and malignant. In study [32], hybrid deep learning model is used to solve class imbalance problem. Patches of size 224 × 224 × 3 are provided to pre-trained ResNet50 for feature extraction. Feature vector of each patch of an image is inputted to kernelized weighted extreme learning machine (KWELM) which provides weights to each feature of all instances. Instances of minority class are assigned high weights and and low weights are assigned to majority class instances. This model achieved accuracy up to 90.02%. Nucleus guided transfer learning framework is proposed in study [33] where five different pre-trained models: ResNet-18, ResNet-50, ResNet-101, GoogleNet, and AlexNet are used for feature extraction. After feature extraction, svm classifier is used at the end of each feature extractor. Output of all svm models is fused on belief based to get single result. This framework achieved 96.91% accuracy, 96.18% specificity, 97.28% sensitivity.

Several approaches including [34], [35], [36], [37], and [38] are introduced for classification of histopathology images based on feature fusion and decision fusion. Multiscale CNNs-based EMS-Net [34] is proposed for the classification of histopathology images. In this approach, input images are divided into three scales; images of original scale, resized images of 448 × 336 and 296 × 224 pixels dimensions. After multi-scaling, patches of 224 × 224 pixels are extracted from all three scales, separately. Pre-trained ResNet-152, DenseNet-161, and ResNet-101 are used to extract features from the input patches. It used both feature fusion and decisoin fusion based approaches. This model obtained 91.75% offline images accuracy in the five-fold cross validation and for online dataset it has 90.00% accuracy. Performance of the model remained low on patches that were extracted from large-scale images. In study [35], patches of size (128 × 128) and (512 × 512) are generated for cell and tissue level features extraction, respectively. ResNet50 models are trained on both types of patches in parallel. P-norm pooling is used to collect extracted 2048-dimensional features group. Extracted features from both types of patches are used for the training of svm classifier. This model achieved 88.89% accuracy on the test dataset. In another study [36], pre-trained VGG16 model is used to classify breast cancer histopathology images. Part-level and whole-level input images are used to extract foreground and background features. In part-level block, patches of 224 × 224 are used while in whole-level block, images are initially resized to 224 × 224 and then used as input. Whole-level inputs are used to extract features
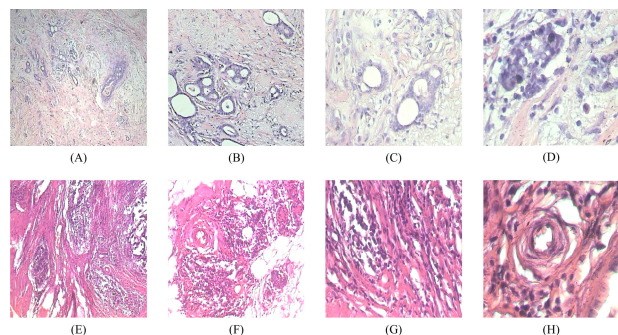
FIGURE 1. (A, B, C, D) are benign class images having magnification levels (40×, 100×, 200×, 400×) respectively and (E, F, G, H) are malignant class images having magnification levels (40×, 100×, 200×, 400×) respectively.

missed by the part-level inputs. SVM classifier is used to classify the input images into four classes. Model achieved 92.2% accuracy on BACH dataset. Gecer et al. [38] concluded in their study that whole-level approaches are better as compared to patch-level approaches. They identified that patch-based biasness of model may affect its overall performance on the whole image. In study [37], decision fusion technique based on majority voting is used in pretrained multi-CNN models for image classification. Feature fusion is considered a better approach as compared to other fusion approaches. In feature fusion, decision is performed by depending on the rich features of all models which are involved in the decision of classification model. Interdependency of models during feature fusion make it superior over other fusion approaches.

III. DATASET
BreaKHis dataset is used for the validation of the proposed MSF model. BreaKHis dataset having histopathology images is introduced by spanhol [6]. Histopathology images regarding this dataset are collected between January and December 2014 for breast cancer classification. It contains 7909 histopathology images of 82 patients. Out of 82 patients, 24 belongs to benign and 58 have malignant class. BreaKHis dataset has two classes benign and malignant. Both classes are further divided into four sub-classes. Sub-classes of benign are adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma and malignant has ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma sub-classes. These images are taken through Olympus BX-50 microscope with a 3.3× magnification relay lens with SCC–131AN. Images of this dataset have four; 40×, 100×, 200× and 400× magnification levels. Images of this dataset are RGB colored portable network graphics (PNG) images of 24-bit color depth and size 700 × 460 pixels. It is a publicly available dataset [39]. Detailed overview of dataset is shown in Table 1. In Figure 1 represents images of various types at different magnification levels.

IV. PROPOSED MULTI-SCALE FEATURE FUSION (MSF) MODEL
The multi-scale feature fusion (MSF) model is proposed in this study. Histopathology images are complex in structure

TABLE 1. Images distribution by class and magnification level.

Magnification Level	No. of Benign Images	No. of Malignant Images	Total Images
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909

due to low difference between foreground and background. Several other reasons of image complexity are blurry boundaries of nuclei, and occluded nuclei. This complexity increased due to poor staining process of histopathology images. So, transfer learning from one domain to breast histopathology causes domain adaptation issue. In this study, MSF has applied same domain transfer learning to overcome the domain adaptation issue. Contrast limited adaptive histogram equalization (CLAHE) which is a contrast enhancement technique is used as pre-processing step to overcome the complexity of histopathology images. Unavailability of large dataset in breast histopathology is another issue. In the proposed MSF, rotation, flipping, and scaling are used as data augmentation techniques. Disappearing of small objects in deep layers is an issue of deep learning models. Dilated layers in six heterogeneous multi-models are used to retain small objects in deep layers.

Performance of any model depends upon features extraction. A better features extractor helps the classifier to identify true classes of inputs. The proposed MSF model has total six feature extractors which are divided into two blocks. It has utilized two scales 224 × 224 × 3 and 512 × 512 × 3 as inputs. Feature extractors of the MSF model are divided into two blocks. Block 1 has three feature extractors: modified ResNet101-1, modified EfficientNetB3-1, and modified DenseNet121-1. Each feature extractor in block 1 takes input of 224 × 224 × 3 dimensions. Whereas in block 2, modified ResNet101-2, modified EfficientNetB3-2, and modified DenseNet121-2 take input of 512 × 512 × 3 dimensions for features extraction. The purpose of using multi-scale inputs is to extract those features which are not possible for a stand-alone model to extract using single input scale. Size of input images affects the performance of models. Similarly, a lot of improvement is seen in hybrid and ensemble models as compared to stand-alone feature extractor. The MSF is capable enough to cope with the issue of image complexity. Finally, extracted features are fused and reduced in feature concatenation and convolution layers, respectively. Fused features are further passed through three fully connected layers for feature engineering and image patterns recognition. Structure of the proposed MSF is given in Figure 2.

A. DOMAIN ADAPTATION
Transfer learning is used to overcome the unavailability of large dataset. It is also useful for fast training of the model. It is used to achieve better results with less epochs. Transfer learning from other domain to histopathology images cause
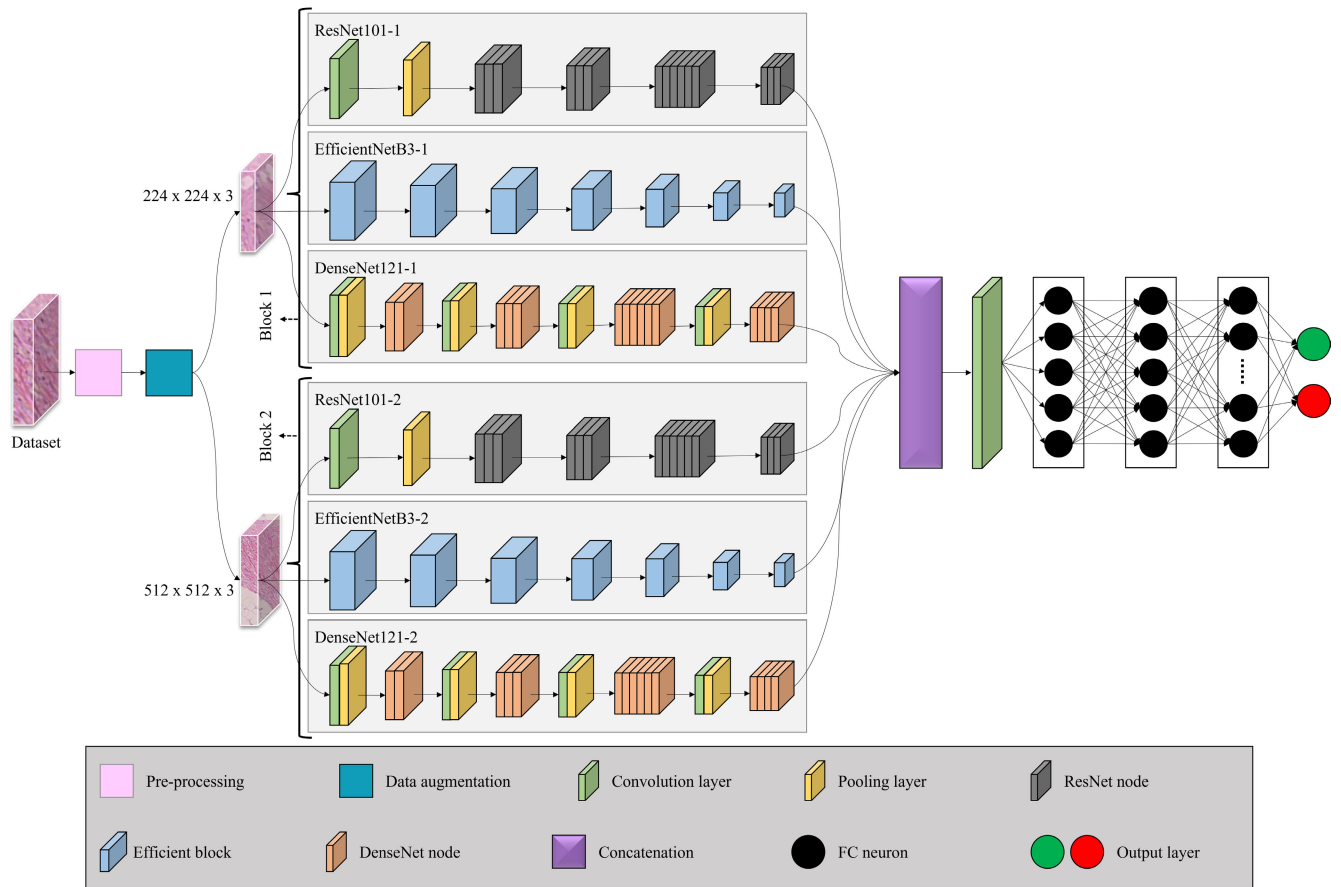
FIGURE 2. The proposed Multi-Scale Feature Fusion (MSF) model for breast cancer classification.

domain adaptation issue. This issue occurs due to differences in between natural images of ImageNet dataset and microscopic images. Natural images have various object types and size while microscopic objects have small size and similar shapes. Therefore, the MSF model has used same domain transfer learning to overcome domain adaptation issue. The proposed model is trained on Patchcamelyon17 dataset from scratch. It is a time taken task however, it provided significant improvement in performance. Model is trained on Patchcamelyon17 dataset for 100 epochs to achieve weights of same domain. Patchcamelyon17 is a binary class breast cancer classification dataset. Pre-trained weights of same domain are further used for the training of the MSF model.

B. PRE-PROCESSING AND DATA AUGMENTATION
Contrast enhancement is one of the important tasks for image quality enhancement. Contrast in histopathology images is mostly low between foreground and background. It could be due to poor staining process or weak quality of microscope for image acquisition. Histopathology images are complex, blurry, and occluded. The proposed MSF model is experimented over histogram equalization (HE), adaptive histogram equalization (AHE) and clahe as image enhancement techniques. HE and AHE works on global features by applying

changes in images using global window. Clahe is a small local sliding window-based image enhancement technique. It provided satisfactory results as compared to other experimented techniques. The principle focus of contrast enhancement is to increase contrast between nuclei and background. Therefore, clahe is adopted as pre-processing step for image quality enhancement. Clahe is the alternative of AHE. Contrast in clahe is limited due to small size of window. In clahe, threshold and tile/window size are two important parameters. Threshold is used to limit the contrast. In the proposed MSF model, threshold of three is used. Tile size is another parameter of clahe which determines the size of patch which is used for contrast enhancement in an operation. The MSF model used (8, 8) sized window for contrast enhancement. It use small window for contrast enhancement therefor it is a limited contrast enhancement technique. It is useful for images having small objects in them.

To overcome the unavailability of large dataset, two techniques are used. In the first technique, weights of same domain are achieved for same domain transfer learning, and in second technique data is augmented. The proposed MSF model has six sub-models therefore large data helped it for better performance. So, in this study, conventional data augmentation techniques are used. Rotation at 30, horizontal and

vertical flips, and scaling from 1.1 to 1.2 are used. These data augmented techniques increased the BreaKHis dataset up to 12 times of the standard dataset.

C. FEATURE EXTRACTION

In image classification, patch-based and whole-image based inputs are mostly used. Resizing of whole image has advantages over patch-based approach. In EMS-Net [34], patches from the whole image as well as from resized image are used. However, performance of EMS-Net remained worst on patches extracted from the large-scale image. Sitaula et al. [36] concluded that whole-image based inputs are used to extract features that are missed by the patch-based inputs. Gecer et al. [38] also considered whole-image based approaches as better over patch-based approaches by concluding that patch-based biasness of model may affect its overall performance on the whole image. Therefore, in the proposed MSF model resizing of whole images is used instead of patch-based approach as input for feature extraction.

Feature extraction is one of the factors that affects the performance of models. Better feature extraction techniques lead the models toward higher performance. The proposed MSF model has two feature extraction blocks. Both blocks have different input scales. There are three heterogeneous feature extractors in each block. All three feature extractors in block one take input scale $224 \times 224 \times 3$ whereas $512 \times 512 \times 3$ is the input scale for block 2. Multi-scale and multi-model feature fusion techniques are helpful to extract variety of useful features. All feature extractors are modified forms of ResNet101, EfficientNetB3, and DenseNet121.

1) MODIFIED ResNet101

Deep models are considered best models for deep feature extraction, however, plain convolutional neural networks face vanishing gradient which degrades performance of deep networks. Skip connections in ResNet architecture [40] are introduced as shown in Figure 3 to overcome the issue of vanishing gradient. Stacking feature maps of identity block 'x' through skip connections in ResNet model makes it different from other models. Layers of plain models are computed using equation 1 while ResNet model has different computation and output as shown in equation 2. If any layer at any stage faces performance issue than this skip connection is helpful to overcome that performance degradation issue as it stacks feature maps of previous layer. ResNet uses additive method (+) to add previous layers with the subsequent layers. In equations 1 and 2, 'w' are weights, 'x' is input feature maps, and 'b' is a bias. ResNet model is lightweight model.

$$H(x) = f(wx + b) \tag{1}$$

$$H(x) = f(wx + b) + x \tag{2}$$

ResNet101 has 101 layers and four blocks. The modified ResNet101 has 97 convolution layers, 3 dilated convolution layers and one max pooling layer. In convolution and dilated convolution layers, kernel size of 1×1, 3×3, and 7×7 are used as described in Figure 3. ReLU activation function is utilized in the MSF model. Fully connected and average pooling layers are removed from the modified ResNet101 feature extractor. Use of dilated convolution layers is valuable to preserve the local and the global features of input images. Multiple dilated layers may affect the structure of input images. In the modified ResNet101 model, three dilated convolution layers are used in different blocks. First dilated layer is used in conv2_block1, second is used in conv3_block1, and third is used in conv4_block1. Dilation rate of two is used in all three dilated convolution layers. Dilated layer is used after the decrease in image in three blocks to increase size of objects by preserving their shapes.

2) MODIFIED EfficientNetB3

Various deep learning models are introduced which either follow depth scaling, width scaling, or resolution scaling. The depth scaling process involves in sequentially increasing or decreasing the number of layers in various models. More layers are frequently used to get more complex features. However, performance of plain networks often decreases as their depth is increased. In width scaling, several layers work concurrently in the form of branches. It is beneficial to obtain fine-tuned features with comparably less depth. However, the accuracy of wide networks reaches to saturation as the width increases. The increase in input resolution is referred to as resolution scaling. Literature depicts that accuracy of model increases with resolution. However, after certain resolution, accuracy is unaffected, and larger inputs also prolong the computation time of models. Combination of all scales in a model, solves these issues. EfficientNet is one of the prominent deep learning models. It scales the CNNs in an efficient way by combining three dimensions: depth scaling, width scaling, and resolution scaling. This combination makes the EfficientNet capable to extract complex, and fine-gradient features. So, the performance of EfficientNet improves by balancing the combination of its depth, width, and image resolution.

EfficientNetB3 is modified in the MSF model. Modified EfficientNetB3 is shown in Figure 4. It has total seven blocks. It contains stem, three modules, and addition to concatenate blocks and modules. Stem and modules are further divided into various layers as shown in Figure 5. Stem is a starting point of EfficientNet which takes image as input and resize it into defined size. Input is normalized after rescaling. Zero padding is applied after normalization. After zero padding conv2d, batch normalization, and activation function are used in stem, respectively.

M1, M2, M3 are module 1, module 2, and module 3, respectively. The modified EfficientNetB3 contains one M1, six M2 and nineteen repetitions of M3. M1 and every M2 has M3 as successor. Addition of feature maps is performed after every M3. M1 is the sequence of depth-wise conv2D →batch normalization →and activation function whereas M2 is the sequence of M1 →zero padding →and M1, respectively.

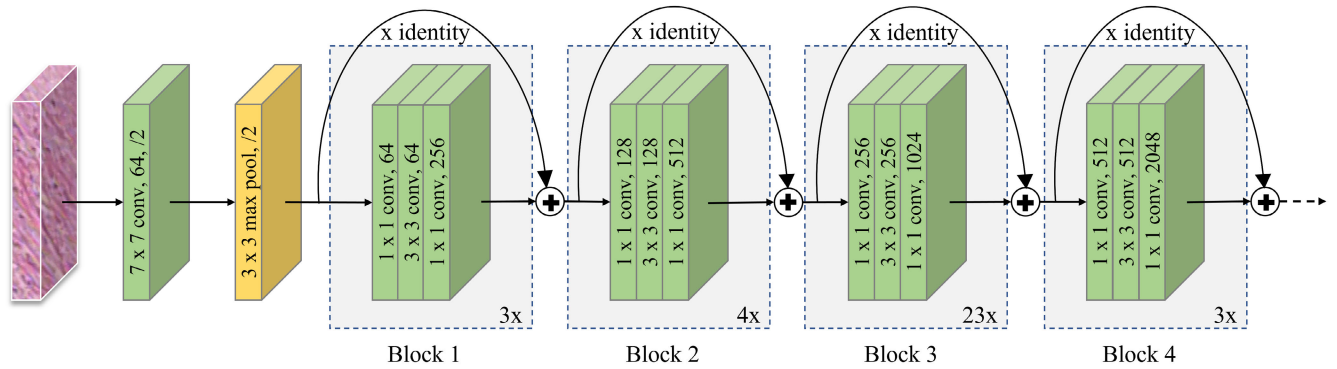**FIGURE 3.** Modified ResNet101.

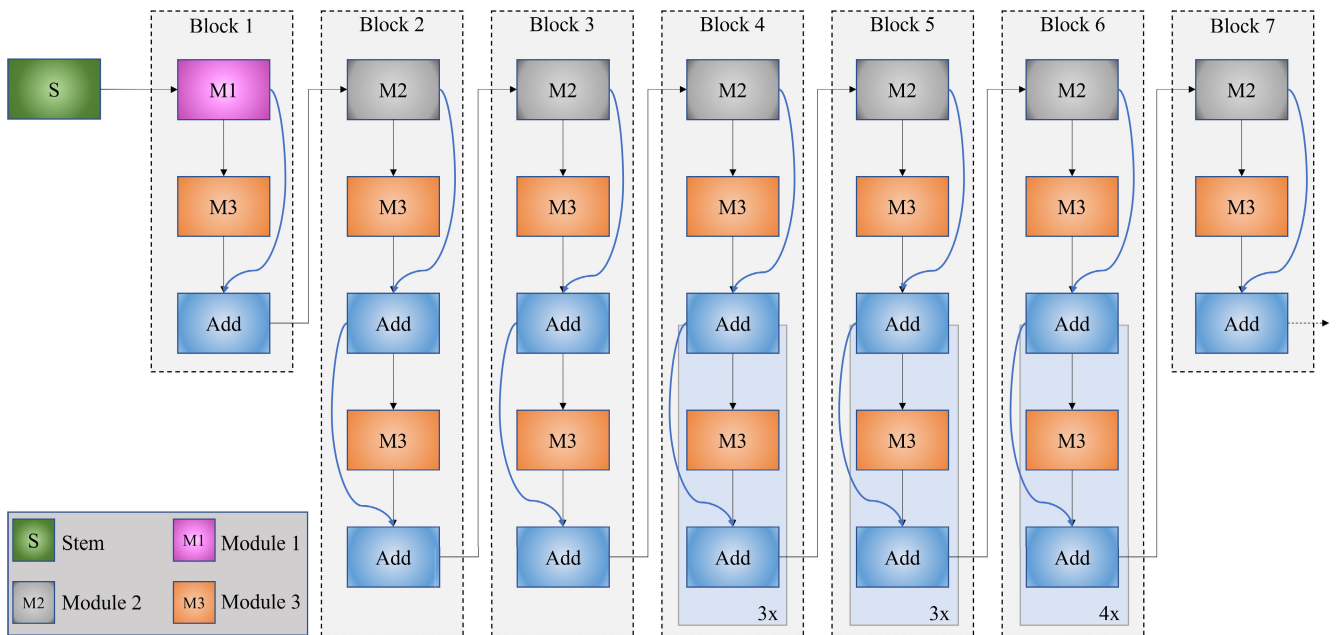**FIGURE 4.** Modified EfficientNetB3.

M3 is the combination of global average pooling →rescaling →conv2D →conv2D except the modified M3 as shown in Figure 5. In the modified EfficientNetB3, total four M3 modules are modified including first M3 of block3, block4, block5 and block6. Modified M3 is the combination of global average pooling →rescaling →dilated conv2D →conv2D.

3) MODIFIED DenseNet121

Densely connected convolutional network [41] is introduced in ILSVRC challenge 2017. DenseNet uses concatenation method (.) to connect output of preceding layers with subsequent layers. In DenseNet, every layer is directly connected to all next layers so each layer in this model has collective knowledge of all previous layers. This connection is helpful to overcome vanishing gradient. Total connections of DenseNet model can be calculated using the equation 3

where L is the number of layers in the model. This connection is only possible if layers have feature maps of same dimensions.

$$TotalConnections = \frac{L(L+1)}{2} \qquad (3)$$

DenseNet model is divided into dense blocks and transition layers. Dense blocks contain multiple convolutional layers with different number of filters but each layer in the block has same dimensions of feature maps. Dense blocks in the model have different dimensions so transition layers are used to downsize the feature maps of previous block to connect with the next blocks. A transition layer consists of batch normalization →activation function →convolution 1 × 1 →drop out layer →and pooling 2 × 2 as shown in Figure 6. This model extracts diversified features as every layer in this model takes feature maps from all previous layers. It is faster
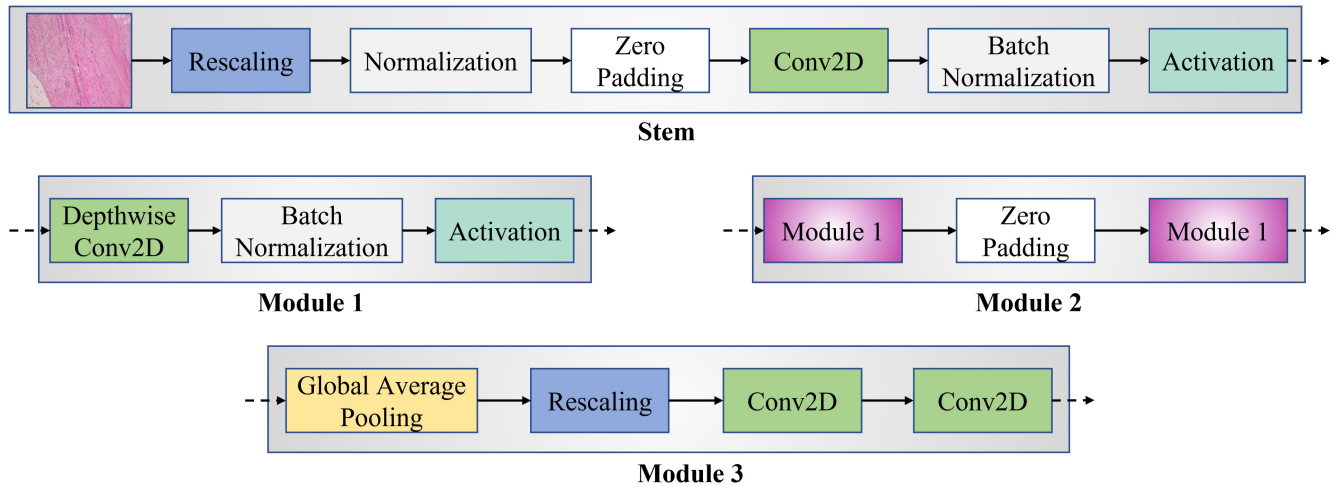
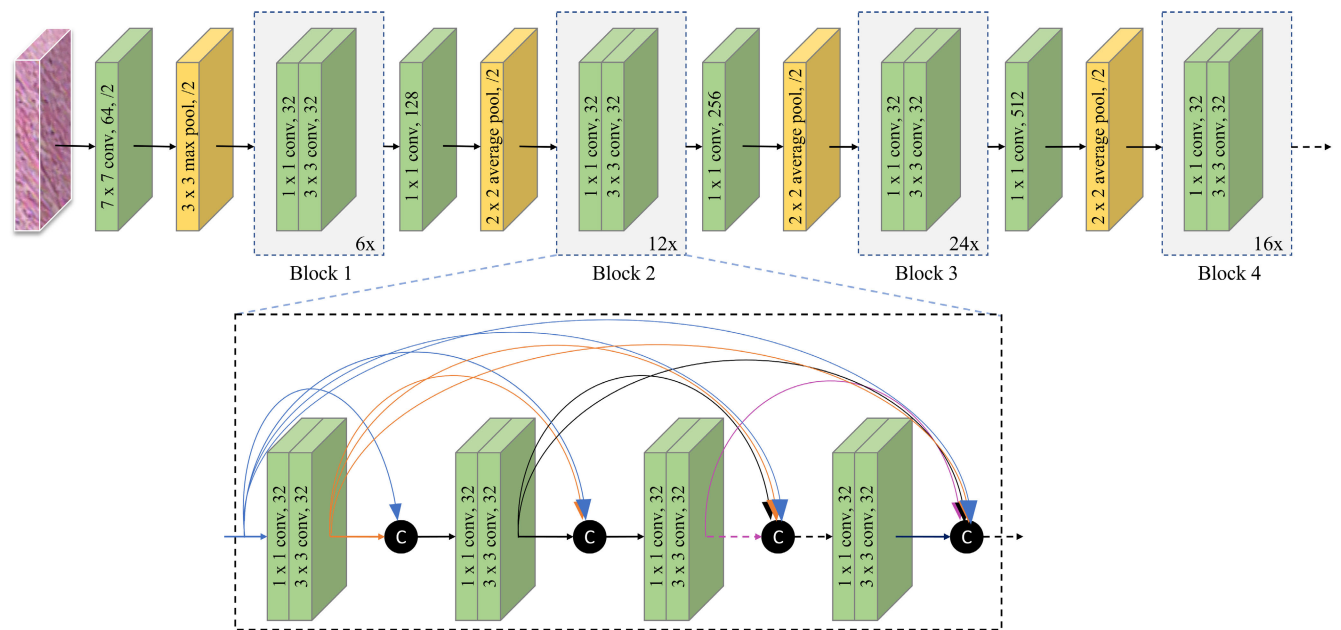**FIGURE 5.** Modules of modified EfficientNetB3.

FIGURE 6. Modified DenseNet121.

and memory efficient due to a smaller number of additional channels.

Modified DenseNet121 has 116 convolution layers, four dilated convolution layers, and four pooling layers. First convolution layer of modified DenseNet121 has 7×7 kernel size with stride of two. All other convolution layers including the dilated convolution layers have stride one and 1×1, 3×3 kernels' size. The modified DenseNet121 has total four blocks. In the modified model, total four dilated layers are added, one in each block. Every block starts with convolution layer having kernel size 1×1. Second layer of each block is modified to dilated convolution layer with kernel size of 3×3 with dilated rate of two. Dilated layers are added

to extract detailed local features. It is helpful to increase small patterns in images by decreasing the probability of disappearing of important information. First pooling layer of modified DenseNet121 is a max pooling layer having kernel size of two and other three layers are average pooling layers with kernel size 2×2 and stride of two.

D. FEATURE FUSION AND CLASSIFICATION

Regarding the fusion, decision fusion and feature fusion can be the good consideration. Decision fusion is late fusion which could be based on average voting, maximum voting, or majority voting whereas feature fusion is mostly used early fusion technique. Effective integration of features in feature

fusion technique makes it superior over decision technique. Feature fusion is concatenation of feature extracted from various models, hence weak features of one model could be covered with the rich features of other models. In decision fusion, all models make independent decision based on their own features. These features could be weak which may affect the overall decision of the model. However, in feature fusion, models make decision based on rich features of all fused models so, interdependency of models in feature fusion makes them superior over the other fusion techniques. Therefore, feature fusion technique is used in the proposed MSF model.

Features from the six sub-models are extracted in the proposed MSF model. Purpose of using heterogeneous multi-models is to extract diverse features for better performance. Features of all six models are fused using concatenation function. Dimensions of features are reduced during concatenation process to remove redundant features. This process is helpful to reduce computational time. Feature maps are further provided to convolution layer that has kernel size 3×3 and stride of two. Output of convolution layer is further provided to sequence of three fully connected layers. These layers are further used to refine the features of multi-models. Finally, SoftMax layer is used to classify input data into two class breast cancer classification. The proposed MSF model is implemented on BreaKHis dataset which represents benign as normal images and malignant as cancerous images. Green circle of Figure 2 represents benign class whereas red circle indicates the malignant class.

V. RESULTS AND DISCUSSION

The MSF model is implemented in python programming language. The proposed model is trained for 100 epochs on BreaKHis dataset. Data is divided into 70%, 20%, and 10% for training, validation, and testing of the proposed model, respectively. Results of the proposed MSF model are discussed and compared with other existing studies.

A. EVALUATION METRICS AND COMPLEXITY OF MODEL

Performance of the MSF model is measured using precision, recall, f-measure and accuracy, evaluation metrics as shown in following equations 4, 5, 6, and 7.

$$\text{Precision} = \frac{T_P}{T_P + F_p} \tag{4}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{5}$$

$$\text{F} - \text{Measure} = \frac{2 * \text{precision} * \text{Recall}}{\text{precision} + \text{Recall}} \tag{6}$$

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_N + T_N} \tag{7}$$

F-measure is calculated on the basis of precision and recall. Precision and recall rely on the values of correctly classified samples. True positive (T_P) are correctly classified samples. False positive (F_P) samples are those samples which are classified as cancerous images while these were non-cancerous. False negative (F_N) are considered as non-cancerous while

these were cancerous. True negative (T_N) are correctly classified samples.

Time Complexity of any model depends on the number of arithmetic operations performed by it for an image during forward pass operation. Input size, image channels, padding, stride, filter size, number of layers, and layer type are the factors that affect the number of operations in convolution, pooling, and fully connected layers. Arithmetic operations of the convolutional layer are calculated using equation 8 as given below.

$$\begin{aligned} \text{conv\_operations} = {} & k(cwh + 1)\left(\frac{M - w + 2p_w}{s_w} + 1\right) \\ & \times \left(\frac{N - h + 2p_h}{s_h} + 1\right) \end{aligned} \tag{8}$$

In all equations 'k' represents the number of biases and 'c' is used for the number of channels of both input image and filters. 'w', and 'h' denote width and height of filters used in different layers, respectively. Width and height of the input image to any layer is represented with 'M', 'N', respectively. Widthwise or horizontal padding and stride are p_w and s_w, respectively. Vertical or heightwise padding is denoted with p_h and stride is denoted with s_h. Number of operations performed by any pooling layer is calculated using equation 9.

$$\begin{aligned} \text{pool\_operations} = {} & c(wh - 1)\left(\frac{M - w + 2p_w}{s_w} + 1\right) \\ & \times \left(\frac{N - h + 2p_h}{s_h} + 1\right) \end{aligned} \tag{9}$$

Operations of fully connected flatten layers are calculated using equations 10.

$$\text{FC\_operations} = k(n + 1) \tag{10}$$

'n' in equation 10 represents the number of input neurons whereas 'k' represents the number of output neurons or bias connections. Overall operations of any model are calculated by adding the floating point operations (multiplication, addition) of all layers. Total operations of the model are calculated by the addition of operations performed in convolutional, pooling, and fully connected (FC) layers with some extra 'X' operations using equation 11. Limited operations are performed in SoftMax and other activation functions. Normally these extra operations are ignored during calculation.

$$\begin{aligned} \text{Total\_operations} = {} & \sum_{i=1}^{m} \text{conv\_operations}_i \\ & + \sum_{j=1}^{n} \text{pool\_operations}_j \\ & + \sum_{l=1}^{o} \text{FC\_operations}_l + X \end{aligned} \tag{11}$$

TABLE 2. Performance of proposed model at different magnification levels.

Magnification level	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
40×	97.85	97.94	97.90	98.05
100×	98.00	98.15	98.08	98.23
200×	97.75	97.81	97.78	97.90
400×	97.40	97.51	97.46	97.60

B. COMPARISON OF RESULTS AT ALL MAGNIFICATION LEVELS

BreaKHis dataset has four magnification levels; 40×, 100×, 200× and 400×. The proposed MSF model is trained, validated, and tested on all magnification levels. Precision, recall, f-measure, and accuracy are used as evaluation metrics. The proposed MSF model achieved 98.20%, 98.60%, 97.70%, and 97.20% validation precision at 40×, 100×, 200×, and 400× magnification levels, respectively as shown in Figure 7(a).

Validation recall of the proposed MSF model is 97.80%, 98.20%, 97.50%, and 97.10% at 40×, 100×, 200×, and 400× magnification levels, respectively which is represented in Figure 7(b). Magnification levels 40×, 100×, 200×, and 400× have validation f-measure of 98%, 98.15%, 97.73%, and 97.25%, respectively as shown in Figure 7(c). The proposed MSF model achieved 98.30%, 98.70%, 97.80% and, 97.30% validation accuracy at 40×, 100×, 200×, and 400× magnification levels, respectively as given in Figure 7(d). The proposed MSF model achieved highest validation accuracy, precision, recall, and f-measure at 100× magnification level. Steep growth in precision, recall, f-measure, and accuracy of the proposed MSF model for all magnification levels is observed till forty-five epochs. After forty-five epochs this increase become smooth till hundred epochs.

The MSF model is tested for all magnification levels. Results of the proposed model are close to each other for all magnification levels, but it is observed that it performed better at 100× magnification level as compared to the other magnification levels. Highest performances of the proposed model are at 100×, 40×, 200×, 400× in a sequence. Performance of the proposed model can be observed from the Table 2. Graphical representation of the proposed model is shown in Figure 8. Results at 100× are better for the proposed model so, results at magnification level 100× are adapted and further used in remaining sections.

C. COMPARISON OF TRAINING AND VALIDATION RESULTS

The proposed MSF model performed better at 100× magnification level. Model is trained for 100 epochs. Training and validation results of the model are observed for all epochs. Precision, recall, f-measure, and accuracy of the MSF model for training and validation is shown in Figure 9. The MSF model achieved up to 98.51%, and 98.20% precision for training and validation data, respectively as shown in Figure 9(a).

TABLE 3. Evaluation results of the MSF model at 100× magnification level.

Magnification level	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
100×	98.00	98.15	98.08	98.23

Similarly, the model achieved 98.40% training and 98.08% validation recall, as elaborated in Figure 9(b). F-measure of the MSF model for training is 98.45% and 98.17% is for testing as shown in Figure 9(c). Figure 9(d) is showing accuracy of the MSF model. It has achieved 99.01% training and 98.70% validation accuracy on BreaKHis dataset. Smooth increase in the training and validation graphs with the change of epochs, shows better tuning of the model.

The MSF is tested on 10% of the augmented dataset. On test data, the model achieved 98.00%, 98.15%, 98.08%, and 98.23% precision, recall, f-measure and accuracy, respectively at 100× magnification level. Test results are explained in Table 3 and graph of test results is presented in Figure 10. Performance of the model at test data is close to training and validation performance. It indicates that model performed well for unknown samples.

VI. ABLATION STUDY

Ablation study of the proposed MSF model is conducted to dive deep insight into the effects caused by variations and components of the model. The proposed model is compared with the existing state-of-the-art studies.

A. STRUCTURAL VARIATIONS OF THE MSF MODEL

The MSF model has experimented through five structural variations during training and validation process. Effects of each variation over proposed model and their results are discussed in this section of ablation study. Variation 1 is the training of model from scratch whereas transfer learning is used in variation 2. The CLAHE, a pre-processing technique is used in variation 3. Same domain transfer learning is used in variation 4 whereas dilation rates are used in variation 5 to extract features of small objects in deep layers.

In first variation, the MSF model is trained and validated on augmented BreaKHis dataset from the scratch. In this variation, pre-trained weights or transfer learning is not utilized. It is comparatively a slow learning processes which can be seen as variation 1 in Figure 11(a,b,c,d). Figure 11(a,b,c,d) has the validation of model in term of precision, recall, f-measure, and accuracy, respectively. In variation 1, the MSF model achieved 92.10%, 92.23%, 92.17%, 92.50% testing precision, recall, f-measure, and accuracy, respectively that is shown in Table 4.

In variation 2, initially, the MSF model is trained on the ImageNet dataset for hundred epochs to obtain pre-trained weights. These pre-trained weights of the model are used as transfer learning for further training of the MSF model at BreaKHis dataset. It is analyzed from various studies that transfer learning tunes model in a better and fast way as
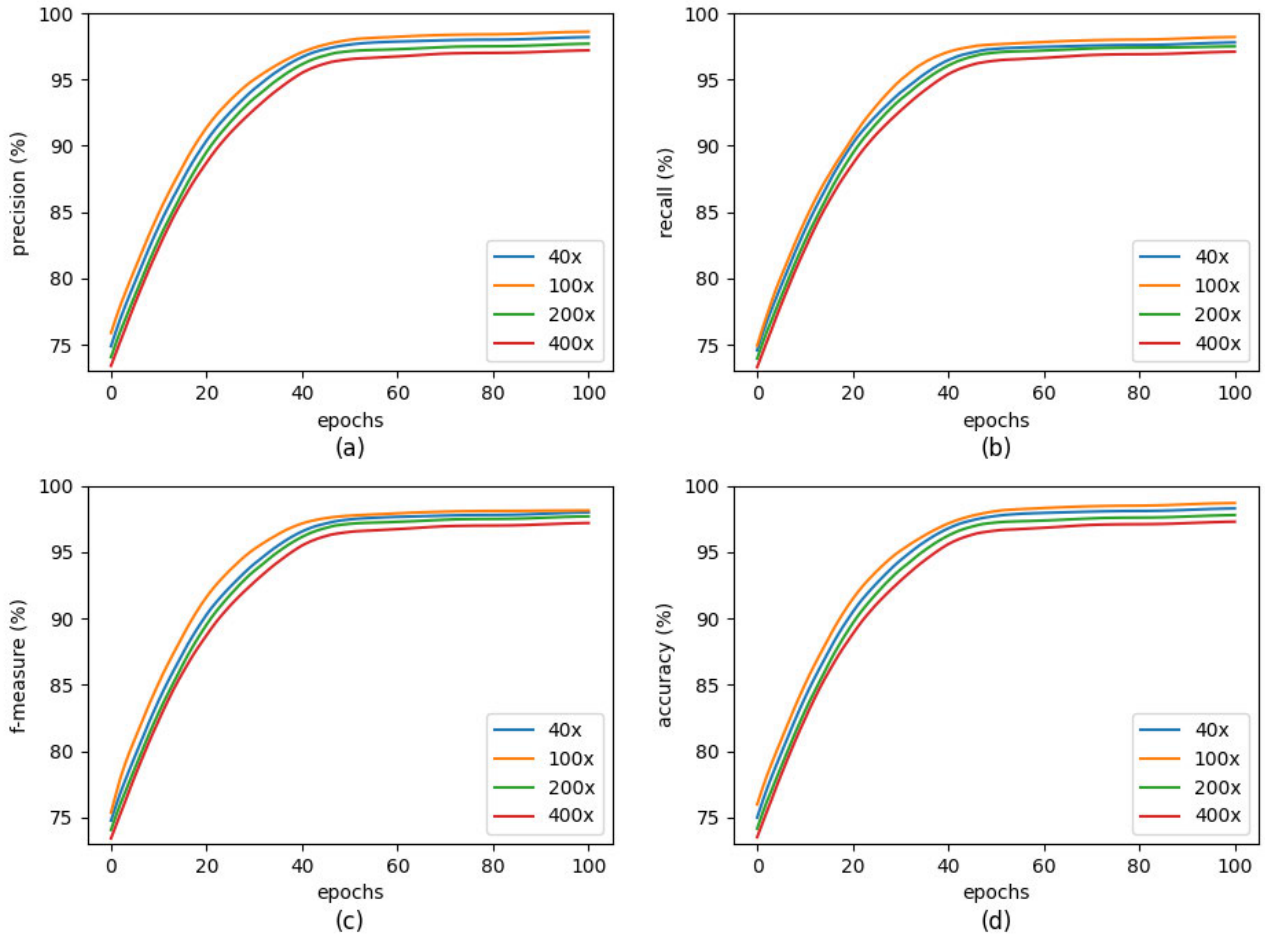
FIGURE 7. Performance of the proposed MSF model at different magnification levels during validation.
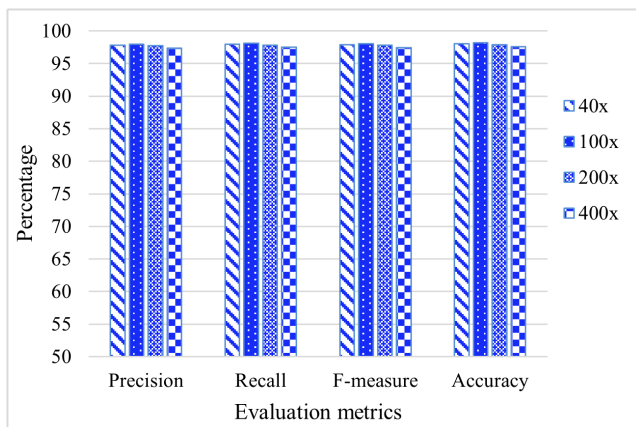

FIGURE 8. Performance of proposed model at different magnification levels.

compared to the training from the scratch. Precision, recall, f-measure, and accuracy of variation 2 during validation of model for 100 epochs is shown in Figure 11(a,b,c,d), respectively. Test results of the MSF model for variation 2 are shown in Table 4. It can be observed that variation 2 of the

model achieved 94.60%, 94.64%, 94.62%, 94.65% testing precision, recall, f-measure, and accuracy, respectively. The model achieved better results at transfer learning as compared to learning from the scratch.

Variation 3 is the advancement of variation 2. In third variation, CLAHE, a pre-processing technique is used on BreaKHis dataset for contrast enhancement and noise removal. During this variation transfer learning is applied on CLAHE-based pre-processed dataset. It improved test accuracy up to 0.78%. Validation results of precision, recall, f-measure and accuracy during variation 3 are plotted in Figure 11(a,b,c,d), respectively. This variation of the model achieved 95.10% precision, 95.30% recall, 95.20% f-measure, and 93.43% accuracy at test dataset as shown in Table 4.

Various studies resulted that different domain transfer learning for complex histopathology images is not the best solution. Structure of microscopic images is different as compared to natural images of ImageNet dataset. To overcome domain adaptation issue, same domain transfer learning technique is adapted in variation 4. Initially, model is trained for 100 epochs at patchcamelyon17 dataset. Pre-trained weights
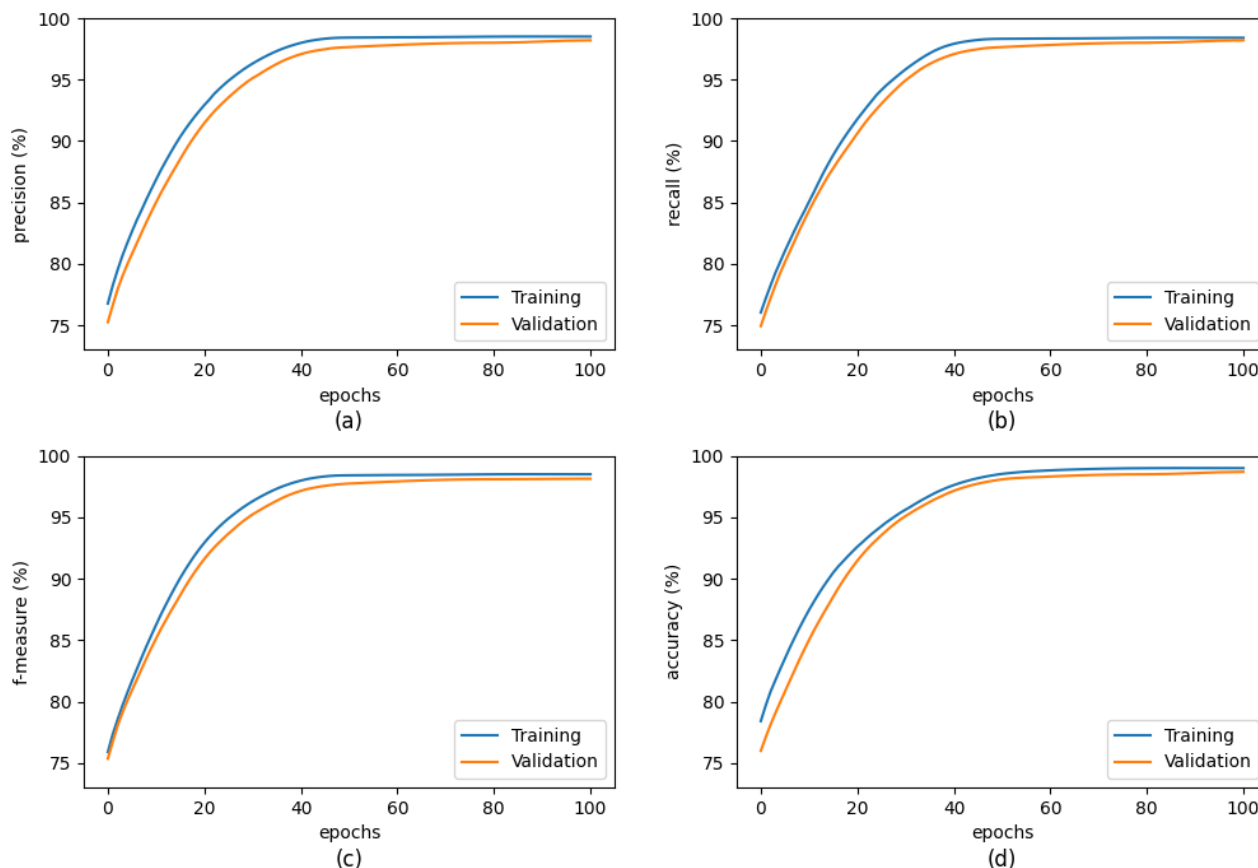
FIGURE 9. Evaluation results of the MSF model at 100× magnification level during validation.
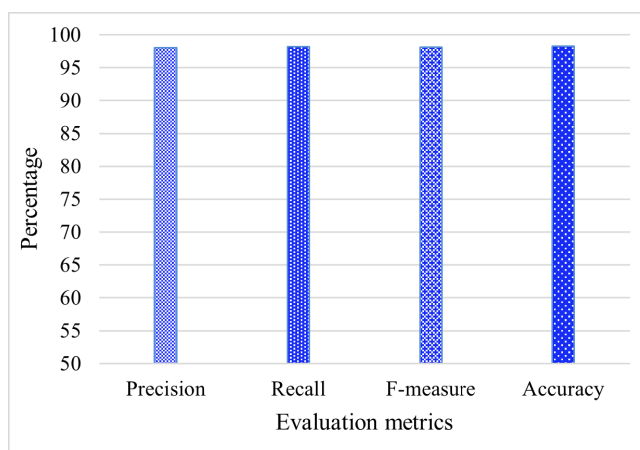

FIGURE 10. Evaluation the MSF model at 100× magnification level.

TABLE 4. Structural variations of the MSF model and their test results.

Model's Variations	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
Variation 1	92.10	92.23	92.17	92.50
Variation 2	94.60	94.64	94.62	94.65
Variation 3	95.10	95.30	95.20	95.43
Variation 4	96.95	97.00	96.98	97.14
Variation 5	98.00	98.15	98.08	98.23

96.98% f-measure, and 97.14% accuracy at test dataset during variation 4 as shown in Table 4.

Variation 5 is the final structure of the proposed MSF model. Results of this proposed variation are also compared in previous sections. It is the advancement of variation 4. In this variation, dilated layers are used to overcome the disappearing of small objects in histopathology images in deep layers. These dilated layers are used in all feature extractors of proposed MSF model. Structure and dilated layers are already explained in the proposed MSF model section. In variation 5, same domain transfer learning-based MSF model with dilated layers is applied on CLAHE-based pre-processed BreaKHis dataset. The MSF model improved test accuracy up to 1.63% in this variation. Validation results of precision, recall, f-measure and accuracy of variation 5 are shown in

of the model at patchcamelyon17 are further used for training of the model at BreaKHis dataset. Both patchcamelyon17 and BreaKHis are histopathology datasets of breast cancer. In this variation, the model improved test accuracy up to 1.49%. Validation results of precision, recall, f-measure and accuracy during variation 4 are shown in Figure 11(a,b,c,d), respectively. Model achieved 96.95% precision, 97.00% recall,
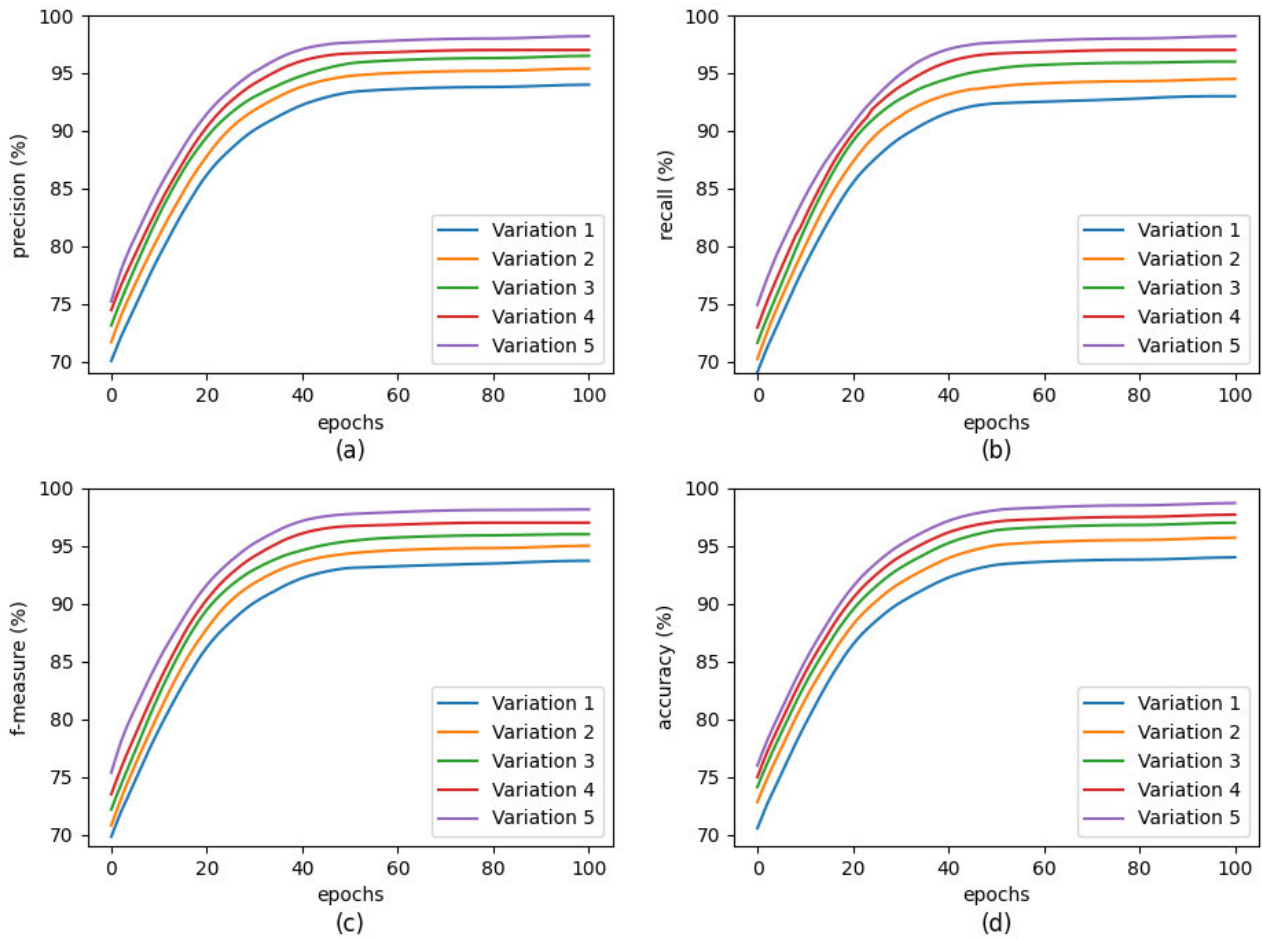
FIGURE 11. Model variations and their performance during validation.

Figure 11(a,b,c,d), respectively. The MSF model achieved 98.00% precision, 98.15% recall, 98.08% f-measure, and 98.23% accuracy at test dataset during variation 5 as shown in Table 4. Overall, the proposed MSF model improved precision 5.90%, recall 5.92%, f-measure 5.91%, and accuracy 5.73% from variation 1 to variation 5 which is the proposed MSF model. Graphical representation of all variations is shown in Figure 12.

B. RESULTS OF ALL COMBINATIONS

In the proposed MSF model, six modified sub-models including ResNet101-1, EfficientNetB3-1, DenseNet121-1, ResNet101-2, EfficientNetB3-2, and DenseNet121-2 are used for feature extraction. Initially, all models are trained on patchcamelyon17 dataset, and their weights are used as same domain transfer learning. ResNet101-1, EfficientNetB3-1, and DenseNet121-1 as a combination make block 1 of the proposed model whereas sub-models: ResNet101-2, EfficientNetB3-2, and DenseNet121-2 make block 2 of the proposed MSF model. The MSF model is the combination of block 1 and block 2. All sub-models of block 1 take $224 \times 224 \times 3$ dimensions input while sub-models of block 2
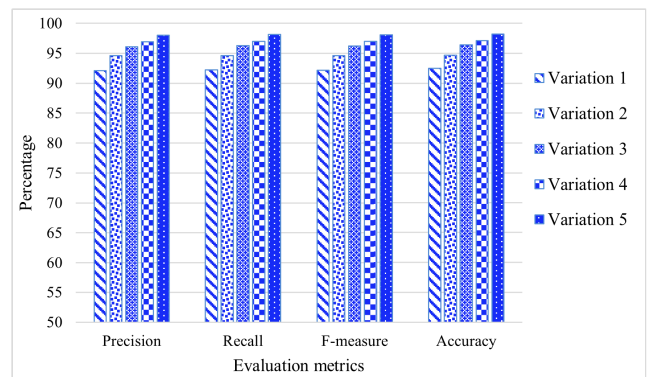

FIGURE 12. Structural variations of the MSF model and their test results.

take input of $512 \times 512 \times 3$ dimensions. In this ablation study, individual results as well as possible combinations of sub-models are described.

Use of multi-scale inputs is helpful to extract various rich features which increased the overall performance of the model. Features fusion extracted from all six components

TABLE 5. Comparison of model with its components.

Model	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
ResNet101-1	91.95	92.05	92.00	92.15
ResNet101-2	92.30	92.46	92.38	92.71
EfficientNetB3-1	90.30	90.35	90.32	90.41
EfficientNetB3-2	90.61	90.70	90.65	90.75
DenseNet121-1	93.88	94.00	93.94	94.13
DenseNet121-2	94.25	94.31	94.28	94.40
Block 1	96.20	96.27	96.24	96.31
Block 2	96.47	96.58	96.53	96.60
Proposed MSF	98.00	98.15	98.08	98.23

TABLE 6. Computational complexity of models.

Model	Parameters (M)	Computational complexity (G) FLOPs
ResNet101-1	44.55	7.85
ResNet101-2	44.55	41.0
EfficientNetB3-1	12.23	0.99
EfficientNetB3-2	12.23	5.18
DenseNet121-1	7.98	2.88
DenseNet121-2	7.98	15.06
Proposed MSF model	129.42	72.96

boosted overall performance of the MSF model as shown in Figure 13(a,b,c,d).

Results during test phase are explained in Table 5. The precision, recall, f-measure, and accuracy of the proposed MSF model are higher as compared to its sub-components and blocks. This result provided two indications. One is that feature fusion of multiple models is better technique as compared to stand-alone feature extractor as feature fusion has better results than stand-alone model. Other indication is that feature extraction from multi-scale models is better as compared to feature extractions from the single scale. Same models at various scales provided different results.

Test accuracies of sub-models ResNet101-1, EfficientNet B3-1, DenseNet121-1, ResNet101-2, EfficientNetB3-2, and DenseNet121-2 are 92.15%, 90.41%, 94.13%, 92.71%, 90.75%, and 94.40%, respectively. Combination of ResNet101-1, EfficientNetB3-1, and DenseNet121-1 has provided 96.31% test accuracy whereas combination of ResNet101-2, EfficientNetB3-2, and DenseNet121-2 has test accuracy of 96.60%. Combination of all six models has test accuracy of 98.23% which is better than its block 1 and block 2 which have 96.31%, and 96.60% accuracy. Similarly, precision, recall, and f-measure of all possible combinations are shown in Figure 14. It is observed that feature fusion of multi-models is helpful to boost the performance of the model by combining their rich features. In feature fusion, limitations of one model could be resolved by overcoming with the strengths of other models

The MSF model is the combination of six sub-models. Complexity of all six sub-models including their parameters is shown in Table 6. Execution time of models depend on hardware resources which are used for computation. According to the rules of algorithm's complexity, max of the parallel threads is taken as time complexity of models using equation 12. So, ResNet101-2 sub-model with $512 \times 512 \times 3$ sized input images has the highest complexity. In best case, complexity of the MSF model is the computational complexity of ResNet101-2 which is 41 giga (G) operations.

$$
\begin{aligned}
\text{MSF\_model}_{\text{time\_complexity}} \\
= max(\text{ResNet101}_1, \\
\times \text{ResNet101}_2, \text{DenseNet121}_1, \text{DenseNet121}_2, \\
\times \text{EffecientNetB3}_1, \text{EffecientNetB3}_2) + X \quad (12)
\end{aligned}
$$

The MSF model has a total computational complexity of 72.96 G operations which is the combination of operations of all sub-models, calculated using equation 13. Computational time of 72.96 G operations is dependent on clock speed of processing hardware. High value of floating-point operations per second (FLOPS) takes less time to compute the operations whereas less value of FLOPS takes higher computational time. Space complexity of models depends on the number of parameters. High number of parameters takes higher space as compared to a smaller number of parameters.

$$
\begin{aligned}
\text{MSF\_model}_{\text{computational\_cost}} \\
= \text{ResNet101}_1 + \text{ResNet101}_2 + \text{DenseNet121}_1 \\
+ \text{DenseNet121}_2 + \text{EffecientNetB3}_1 \\
+ \text{EffecientNetB3}_2) + X \quad (13)
\end{aligned}
$$

C. COMPARISON WITH EXISTING STUDIES

The MSF model has focus to extract rich effective features by using domain adaptive multi-scale heterogeneous feature fusion technique. In this ablation study, results of the proposed MSF model are compared with state-of-the-art ResHist [7], Kernelized Weighted Model (KWM) [32], multi-instance support vector machine (pdMISVM) [42], Resolution Adaptive Network (RAN) [43], MA-MIDN [44], DBLCNN [45], GLCM [46], and AE + Siamese Network (AE+SN) [47] studies which are implemented on the same dataset. The proposed MSF model has 72.96 G operations. On the other hand, MA-MIDN [44] used 4.12 G operations whereas 11.58 G and 65.92 G operations are used in ResHist [7] and Kernelized Weighted Model (KWM) [32], respectively. In GLCM [46], 4.37 G and in DBLCNN [45] 2.36 G operations are performed during feature extraction. Resolution Adaptive Network (RAN) [43] and autoencoder based Siamese networks [47] used resolution adaptive, multiple feature extractors having higher operations as compared to other studies. It is observed from the comparative studies that models having rich feature extractors, achieved better performance. The proposed MSF model has a higher number of operations due to multiple parallel feature extractors which made the model to extract rich features.

It is observed that the proposed MSF model outperformed the existing state-of-the-art models in precision, recall, f-measure, and accuracy. Precision of the proposed model is second highest as compared to state-of-the are models.
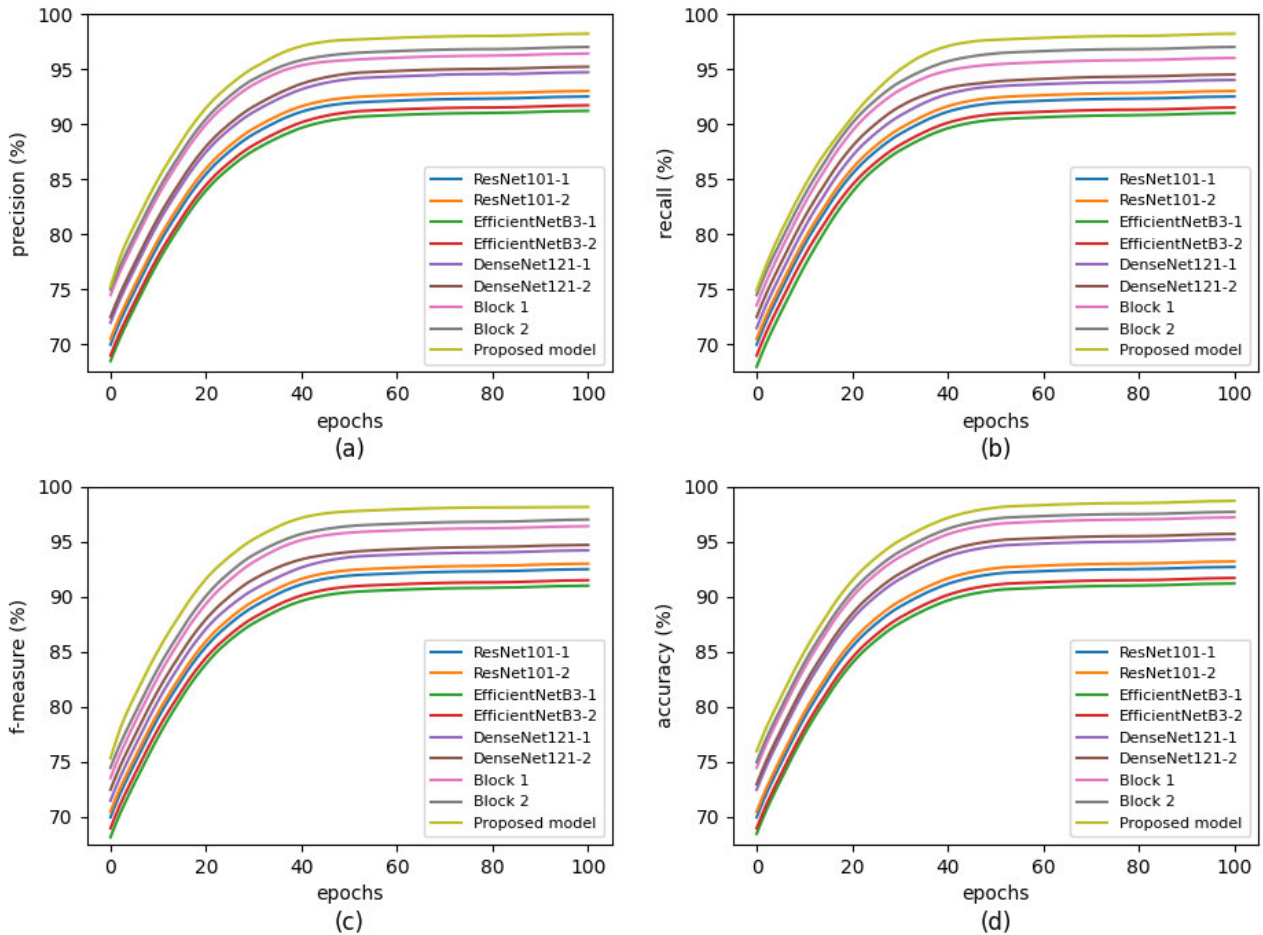
FIGURE 13. Comparison of model with its components and validation results.

TABLE 7. Comparison of the proposed model with existing studies.

Existing studies	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
ResHist [7]	91.87	89.54	90.58	88.15
Kernelized Weighted Model (KWM) [32]	94.71	85.66	89.90	87.14
pdMISVM [42]	87.30	94.40	91.90	88.30
Resolution Adaptive Network (RAN) [43]	98.12	97.80	98.02	97.91
MA-MIDN [44]	96.08	95.51	95.69	95.65
DBLCNN [45]	96.21	95.43	95.97	95.65
GLCM [46]	97.47	95.99	96.71	95.23
AE + Siamese Network (AE+SN) [47]	95.70	98.70	97.20	96.10
Proposed MSF Model	98.00	98.15	98.08	98.23

The MSF model has precision 98.00% whereas highest precision of existing study, RAN [43] is 98.12%. pdMISVM [42] has the lowest precision which is 87.30%. Precision of the ResHist, KWM, MA-MIDN, DBLCNN, GLCM and AE+SN is 91.87%, 94.71%, 96.08%, 96.21%, 97.47%, and 95.70%, respectively. A framework which is designed with the combination of autoencoder and siamese framework [47] has the highest recall value 98.7%. Recall of the proposed model is second highest which is 98.15% whereas kernelized weighted extreme learning model [32] has the lowest 85.66% recall. Recall of other ablative studies including the ResHist,

pdMISVM, RAN, MA-MIDN, DBLCNN, and GLCM is 89.54%, 94.40%, 97.80%, 95.51%, 95.43%, and 95.99%, respectively.

The MSF model has f-measure of 98.07% which is highest as compared to other models. Kernelized weighted extreme learning model [32] has f-measure of 89.90% which is the lowest value as compared to other studies. Accuracy of the proposed model surpassed the existing studies. It has accuracy 98.23% whereas its closest competitor resolution adaptive network [43] has accuracy 97.91%. Kernelized weighted extreme learning model [32] has the lowest accuracy of
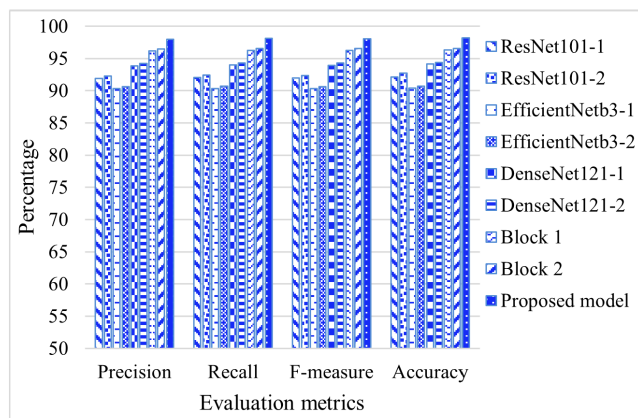
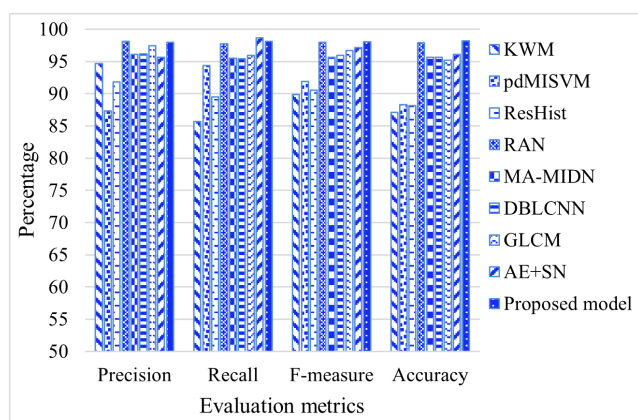**FIGURE 14.** Comparison of model with its components.

FIGURE 15. Comparison of the proposed model with existing studies.

87.14% on BreaKHis dataset. Precision, recall, f-measure, and accuracy of models is shown in Table 7 whereas graphical comparison of evaluation metrics of models is shown in Figure 15. It can be observed that the proposed model has smooth performance on breast histopathology dataset. All evaluation metrics have results close to each other whereas some of the other models perform well for one evaluation metric but not for the others.

The proposed MSF model performed better as compared to state-of-the-art classification models. Same domain transfer learning is one of the reasons that made it prominent. Transfer learning from natural images to microscopic, low contrast and complex images is one of the reasons for low performance of models. Multi-scale inputs provide advantages to models to extract rich features at different resolutions so, inputs of two different scales are used in the proposed model. It is concluded by numerous studies that plain deep learning models provide better results however, they face vanishing gradient problems when depth of plain deep learning models is increased. To avoid vanishing gradients, branch based parallel deep learning models having direct connections to their previous blocks are used in this study. This approach made the proposed model superior as compared to state-of-the-art models. Another advantage of the proposed model

over existing studies is the use of dilated layers. These layers are used in feature extractors to retain small objects in deep layers to extract their global features.

VII. CONCLUSION

Early cancer classification is one of the important tasks to save human lives. In this study, multi-scale feature fusion model based on domain adaptation is introduced for breast cancer classification. Various studies concluded that transfer learning from natural images to microscopic images cause high misclassification rate. In this study, the model is trained on patchcamelyon17 dataset. It is a microscopic dataset of breast cancer. The pre-trained weights are further used for same domain transfer learning of the MSF model. Same domain transfer learning is used to overcome the issue of domain adaptation in medical images due to complexity of histopathology images. The proposed MSF model is the combination of two blocks which are further divided into six different sub-models. These sub-models took multi-scale input images to extract various features. These features are fused, and redundant data is removed. Dilated layers are used in feature extractors to overcome the disappearance of small objects. The MSF is trained and validated on clahe-based pre-processed and augmented BreaKHis dataset. It achieved 98.00% precision, 98.15% recall, 98.08% f-measure, and 98.23% accuracy on test data. To best of our knowledge, it surpassed state-of-the-art models. In future, same domain multi-scale feature fusion technique could be used for the detection and segmentation of the mitotic nuclei for better identification. This technique could be used for whole slide images.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[2] A. R. Saikia, K. Bora, L. B. Mahanta, and A. K. Das, "Comparative assessment of CNN architectures for classification of breast FNAC images," *Tissue Cell*, vol. 57, pp. 8–14, Apr. 2019.

[3] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Syst. Appl.*, vol. 120, pp. 103–115, Apr. 2019.

[4] F. Azour and A. Boukerche, "An efficient transfer and ensemble learning based computer aided breast abnormality diagnosis system," *IEEE Access*, early access, Jul. 25, 2022, doi: 10.1109/ACCESS.2022.3192857.

[5] R. Yan, F. Ren, Z. Wang, L. Wang, T. Zhang, Y. Liu, and X. Rao, "Breast cancer histopathological image classification using a hybrid deep neural network," *Methods*, vol. 173, pp. 52–60, Feb. 2019.

[6] 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.

[7] M. Gour, S. Jain, and T. S. Kumar, "Residual learning based CNN for breast cancer histopathological image classification," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 621–635, Sep. 2020.

[8] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2440–2445.

[9] S. Saxena, S. Shukla, and M. Gyanchandani, "Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 577–591, Sep. 2020.

[10] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, p. 221, Jun. 2017.

[11] V. I. Iglovikov, A. Rakhlin, A. A. Kalinin, and A. A. Shvets, "Paediatric bone age assessment using deep convolutional neural networks," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 300–308.

[12] A. Rakhlin, "Diabetic Retinopathy detection through integration of deep learning classification framework," *BioRxiv*, Jan. 2018, Art. no. 225508, doi: 10.1101/225508.

[13] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: A survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022.

[14] N. Nasrullah, J. Sang, M. S. Alam, M. Mateen, B. Cai, and H. Hu, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, Aug. 2019.

[15] M. Nilashi, O. Ibrahim, H. Ahmadi, and L. Shahmoradi, "A knowledge-based system for breast cancer classification using fuzzy logic method," *Telematics Informat.*, vol. 34, no. 4, pp. 133–144, Jul. 2017.

[16] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[17] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, Jan. 2019.

[18] Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," *Expert Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112951.

[19] M. Fujita et al., "Classification of primary liver cancer with immuno-suppression mechanisms and correlation with genomic alterations," *eBioMedicine* vol. 53, Mar. 2020, Art. no. 102659.

[20] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, Dec. 2018.

[21] N. Maroof, A. Khan, S. A. Qureshi, A. U. Rehman, R. K. Khalil, and S.-O. Shim, "Mitosis detection in breast cancer histopathology images using hybrid feature space," *Photodiagnosis Photodyn. Therapy*, vol. 31, Sep. 2020, Art. no. 101885.

[22] K. S. Beevi, M. S. Nair, and G. R. Bindu, "Automatic mitosis detection in breast histopathology images using convolutional neural network based deep transfer learning," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 214–223, Jan. 2019.

[23] M. Z. Alom, T. Aspiras, T. M. Taha, T. J. Bowen, and V. K. Asari, "MitosisNet: End-to-end mitotic cell detection by multi-task learning," *IEEE Access*, vol. 8, pp. 68695–68710, 2020.

[24] A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangaiah, H. Chauhan, S. Shrivastava, and R. K. Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Inf. Sci.*, vol. 508, pp. 405–421, Jan. 2020.

[25] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2560–2567.

[26] M. B. H. Thuy and V. T. Hoang, "Fusing of deep learning, transfer learning and GAN for breast cancer histopathological image classification," in *Proc. Int. Conf. Comput. Sci., Appl. Math. Appl.* Cham, Switzerland: Springer, 2019, pp. 255–266.

[27] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proc. Int. Conf. image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 737–744.

[28] L. Alzubaidi, O. Al-Shamma, M. A. Fadhel, L. Farhan, J. Zhang, and Y. Duan, "Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model," *Electronics*, vol. 9, no. 3, p. 445, Mar. 2020.

[29] H. Yao, X. Zhang, X. Zhou, and S. Liu, "Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification," *Cancers*, vol. 11, no. 12, p. 1901, Nov. 2019.

[30] P. Wang, J. Wang, Y. Li, P. Li, L. Li, and M. Jiang, "Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102341.

[31] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognit. Lett.*, vol. 125, pp. 1–6, Jul. 2019.

[32] S. Saxena, S. Shukla, and M. Gyanchandani, "Breast cancer histopathology image classification using kernelized weighted extreme learning machine," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 1, pp. 168–179, Mar. 2021.

[33] K. George, S. Faziludeen, P. Sankaran, and P. Joseph, "Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion," *Comput. Biol. Med.*, vol. 124, Sep. 2020, Art. no. 103954.

[34] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images," *Neurocomputing*, vol. 366, pp. 46–53, Nov. 2019.

[35] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019.

[36] C. Sitaula and S. Aryal, "Fusion of whole and part features for the classification of histopathological image of breast tissue," *Health Inf. Sci. Syst.*, vol. 8, no. 1, pp. 1–12, Dec. 2020.

[37] K. Das, S. P. K. Karri, A. Guha Roy, J. Chatterjee, and D. Sheet, "Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 1024–1027.

[38] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," *Pattern Recognit.* vol. 84, pp. 345–356, Dec. 2018.

[39] (2015). *BreakHis: Breast Cancer Histopathological Database BreakHis*. [Online]. Available: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 4700–4708.

[42] H. Seo, L. Brand, L. S. Barco, and H. Wang, "Scaling multi-instance support vector machine to breast cancer detection on the BreaKHis dataset," *Bioinformatics*, vol. 38, pp. 92–100, Jun. 2022.

[43] Y. Zhou, C. Zhang, and S. Gao, "Breast cancer classification from histopathological images using resolution adaptive network," *IEEE Access*, vol. 10, pp. 35977–35991, 2022.

[44] G. Li, C. Li, G. Wu, D. Ji, and H. Zhang, "Multi-view attention-guided multiple instance detection network for interpretable breast cancer histopathological image diagnosis," *IEEE Access*, vol. 9, pp. 79671–79684, 2021.

[45] C. Wang, W. Gong, J. Cheng, and Y. Qian, "DBLCNN: Dependency-based lightweight convolutional neural network for multi-classification of breast histopathology images," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103451.

[46] Y. Hao, L. Zhang, S. Qiao, Y. Bai, R. Cheng, H. Xue, Y. Hou, W. Zhang, and G. Zhang, "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix," *PLoS ONE*, vol. 17, no. 5, May 2022, Art. no. e0267955.

[47] M. Liu, Y. He, M. Wu, and C. Zeng, "Breast histopathological image classification method based on autoencoder and Siamese framework," *Information*, vol. 13, no. 3, p. 107, Feb. 2022.

HAMEED ULLAH KHAN received the bachelor's degree from the University of Bradford, U.K., in 2012, and the master's degree in computer science from the University of Lahore, Islamabad, Pakistan, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad. He has more than five years teaching and industrial experience. He is actively working in the areas of machine learning, deep learning, medical imaging, 2D/3D image processing, video processing, computer vision, and artificial intelligence.

BASIT RAZA received the master's degree in computer science from the University of Central Punjab, Lahore, Pakistan, and the Ph.D. degree in computer science from International Islamic University Islamabad and University Technology Malaysia, in 2014. He is currently an Associate Professor at the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan. His research interests include database management systems, data mining, data warehousing, medical imaging, machine learning, deep learning, and artificial intelligence. He has authored several articles in refereed journals. He has been serving as a Reviewer for prestigious journals, such as *Applied Soft Computing*, *Swarm and Evolutionary Computation*, *Swarm Intelligence*, *Applied Intelligence*, IEEE Access, and *Future Generation Computer Systems*.

HABIB SHAH graduated the Ph.D. degree from the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, in 2013. He is an Assistant Professor and the Head of the Research Unit, College of Computer Science, King Khalid University, Saudi Arabia. His research interests include artificial intelligence, learning algorithms, data mining techniques, the IoT, time series analysis, and optimization. He has successfully published more than 50 articles in various international SCI and Scopus journals and conference proceedings. He is an editorial board guest editor and act as a reviewer for various journals and conferences. He has also served as a program committee member and a co-organizer for numerous international conferences/workshops. He is currently working on three research projects of KKU and KSA.

ABDUL WAHEED received the master's and Ph.D. degrees in computer science from the Department of Information Technology, Hazara University Mansehra, in 2014 and 2021, respectively, and the Ph.D. research from NetLab-INMC under the School of Electrical and Computer Engineering (ECE), Seoul National University (SNU), South Korea, in 2019, under the HEC Research Program. He is a member of the Crypto-Net Research Group, Hazara University. He serves as an Assistant Professor of computer science at the Department of Computer Science, Women University Swabi, Pakistan. Before this, he served as an Assistant Professor at the Department of Computer Sciences and the Dean of the Faculty of Engineering and Information Technology (FEIT), Northern University, Nowshera, Pakistan. He has numerous publications in journals and international conferences. His research interests include information security, secure and smart cryptography, heterogeneous communications within the IoT, mobile Ad hoc networks (MANETs), wireless sensor networks (WSNs) security, and fuzzy logic-based decision-making theory.

● ● ●