RESEARCH ARTICLE

LMHistNet: Levenberg–Marquardt Based Deep Neural Network for Classification of Breast Cancer Histopathological Images

SOUMYA SARA KOSHY AND **L. JANI ANBARASI**

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India

Corresponding author: L. Jani Anbarasi (janianbarasi.l@vit.ac.in)

ABSTRACT Breast cancer is a deadly disease commonly affecting women. One method to avoid death from breast cancer is to obtain a diagnosis early. Breast cancer detection is a significant area that benefits from the technological advancements in artificial intelligence. A deep neural network architecture for the classification of microscopic images of breast tumor tissue acquired using excisional biopsy has been proposed. A hybrid convolutional neural network model with asymmetric convolutions and Levenberg–Marquardt optimization named as LMHistNet is used for the classification of breast cancer images into binary as well as eight subclasses. Convolution block attention module is incorporated for adaptive feature refining. Convergence time has been significantly reduced by normalizing the input features on models using batch normalization. Training is improved by reducing the internal covariance shift. Hinge loss function is used for better convergence. Magnification-dependent as well as independent binary and eight class classifications are efficiently performed by extracting diversified features from histopathological images. 7909 histopathological images of which 2480 were images of benign (normal) and the remaining images of malignant patients were used for the study. Further, performance evaluation on the dataset using LMHistNet has been analyzed based on the loss and accuracy curve. The experimental findings show that the proposed model obtained good performance over the various magnifications 40X, 100X, 200X and 400X. The accuracy, precision, recall and F1 score obtained are 88, 89, 88 and 88 respectively for multiclass classification into eight subtypes. The accuracy, precision, recall and f1 score are 99 each for binary classification of breast tumor tissues into benign and malignant classes.

INDEX TERMS Breast cancer, convolutional neural networks (CNN), deep learning, histopathological images, Levenberg–Marquardt, transfer learning.

I. INTRODUCTION

Breast cancer is one sort of cancer that significantly influences the mortality rates in woman [1], [2]. Genetic factors such as mutations in BRCA1 or BRCA2 genes and non-genetic factors like age, mammographic density, short breastfeeding periods, obesity, alcohol consumption, etc. [3] and [4] can cause breast cancer. The common symptoms of breast cancer are lumps in the breast, swelling, soreness, pain, nipple discharge other than breast milk etc. [5]. Early

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeswari Sundararajan.

detection can greatly reduce the number of fatalities brought on by breast cancer. Women should carry out monthly Breast Self-Examinations (BSE) to find out abnormal lumps and should undergo proper diagnosis and treatment. Apart from BSE, the popular screening and detection methods to detect breast cancer are Clinical breast Examination (CBE) [6], mammography, Magnetic Resonance Imaging (MRI), Ultrasound, thermography, Breast Specific Gamma Imaging (BSGI) etc. Microwave breast imaging [49] can be used to detect breast cancer in which low power microwaves are transmitted into the breast tissue and the back scattered energy are collected using the antennas positioned around

the breast. Three dimensional images of the scanned breast are generated using the received signals can easily detect the breast cancers at early stage. Focused microwave breast hyperthermia (FMBH) [50] can be effectively used in the detection of breast cancer with low cost and minimum side effect. Compressive thermoacoustic tomography (CTT) is an effective guidance approach for FMBH, to monitor the microwave power distribution in the breast.

Earlier the interpretation of data obtained from the various screening methods was done manually. It was a tough task due to the complexity of data and also required the expertise of medical experts. The manual analysis is time-consuming and the experts may get overloaded, leading to misdiagnosis. Computer-aided diagnostic, or CAD, systems are being used to diagnose breast cancer as a result of technological breakthroughs. Artificial intelligence (AI)-based methods, such as machine learning (ML) [7] and deep learning (DL) techniques, have made it possible to identify breast cancer quickly and effectively nowadays. Machine learning algorithms involves finding out hidden patterns from data which aids the classification and prediction of diseases. In machine learning features are extracted manually which is a cumbersome task and takes time. A number of popular machine learning techniques are used to classify breast cancer images, including random forest (RF), logistic regression, and support vector machines (SVM).

Deep learning techniques with multiple layers are faster than machine learning and automatically extract features. Convolutional neural network [8] also known as convnets is a deep learning model particularly used for computer vision applications. CNN, which consists of convolutional, pooling and fully connected layers, is very useful for detecting and classifying breast cancer images.

A custom convolutional neural network, LMHistNet has been proposed for the binary as well as multiclass classification of breast histopathological images. Binary classification involves classifying into benign and malignant classes. Multiclass classification involves classifying the histopathological images into eight classes of which benign includes adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, and malignant includes mucinous carcinoma, lobular carcinoma, ductual carcinoma, and papillary carcinoma. Both magnification dependent and magnification independent classifications have been performed. Channel block attention module [9] is incorporated with the deep convolutional neural network for the efficacious refinement of intermediate features. The weight updation of the network is done using Levenberg–Marquardt algorithm. Breast Cancer Histopathological Database (BreaKHis) [10], created by collecting microscopic images from 82 patients with four different resolutions has been use to show the efficiency of the proposed LMHistNet.

The following are the key contributions of the proposed LMHistNet model.

- To accurately learn the various classification features, various blocks have been incorporated into the model.

- Convolved Group Normalised block (CGN_B), dual Spatial Factorisation and Grid Reduction ($SFG_{R1} \& SFG_{R2}$) Blocks, Convolution Block Attention (CBA_M) Module, and BN block has been integrated for the effective classification of breast cancer histopathological images.
- The position-invariant local features are normalized using Convolved Group Normalised block (CGN_B) whereas the vanishing gradient problem is solved using the dual Spatial Factorisation and Grid Reduction ($SFG_{R1} \& SFG_{R2}$) Blocks. The channel and spatial features are generated using Convolution Block Attention (CBA_M).
- The training speed of the model is increased using Hinge Loss function.
- Magnification independent and magnification dependent binary as well as multiclass classification has been performed.
- Intermediate feature refinement based on channel and attention modules. Max pooled as well as average pooled features are used for channel attention and the performance is elevated with spatial attention.
- The sum of squared difference between the actual data and the predicted data was reduced by using Levenberg Marquardt optimization.
- To obtain optimal accuracy, hyperparameter tuning was performed.

The remaining section of the paper is structured as follows: Section II provides an overview of the relevant research in the field of breast cancer detection: LMHistNet architecture is detailed in Section III; Experimental results and discussions of custom LMHistNet are described in section IV; conclusion and future work are discussed in section V.

II. RELATED WORK

Enormous methods have been put forward by various authors for the classification of breast cancer images, which is a huge task. Earlier machine learning techniques were used, in which the learning was done automatically without explicit programming. Deep learning is a machine learning approach that estimates the underlying patterns in the analysed data by utilizing multiple layers of neural networks.

A. MACHINE LEARNING APPROACHES

Bacha and Taouali [11] used Radial-Based Function Kernel Extreme Learning Machine algorithm for the diagnosis of Breast cancer, the parameters of which are adjusted using differential evolution algorithm. The best attributes are selected using the Kernel Principal Component Analysis (KPCA) methodology. Experimental findings are validated using Mammographic Image Analysis Society (MIAS) dataset and the wisconsin breast cancer database (WBCD). Dai et al. [12] utilized the ensemble learning property of RF in which the output of multiple weak classifiers were combined to increase the classification accuracy. Extreme Gradient Boosting (EGB) and RF are two ensemble learning algorithms that Kabiraj et al. [13] used to predict breast

cancer. EGB was less accurate than RF, which achieved an accuracy of 74.73%. Omondiagbe et al. [14] reduced the feature dimensionality using Linear Discriminant Analysis and the generated feature set was classified using Artificial Neural Networks (ANN), SVM and Naïve Bayes. The SVM showed better performance compared to the other methods with sensitivity 98.41%, accuracy 98.82%, specificity of 99.07% and area under the receiver operating characteristic curve of 0.9994.

Vijayarajeswari et al. [15] used two dimensional hough transform to extract features from mammogram and the extracted features were classified using SVM achieving an accuracy of 94%. Wu and Hicks [16] classified breast cancer into triple negative and non triple negative based on the extracted features at different threshold levels. The classification was performed using four different models NB, K-nearest neighbour (KNN), Decision tree (DT) and SVM. The SVM classifier obtained better accuracy, sensitivity and specificity compared to the other ML models. Michael et al. [17] performed classification of breast ultrasound images into benign and malignant using five different machine learning classifiers like SVM, KNN, XGBoost, RF and LightGBM where the results are enhanced through optimization using Parzen estimator. With a precision of 100%, accuracy of 99.86% and recall of 99.6%, the Light-GBM classifier outperformed the other models. Ara et al. [18] evaluated the WBCD performance using various machine learning algorithms like RF, KNN, SVM, LR, NB and DT. SVM and RF achieved the top accuracy of 96.5%. Naji et al. [19] analysed five machine learning algorithms including LR, SVM, RF, KNN, DT (C4.5) on WBCD and with an accuracy of 97.2%, SVM was found to perform better than the other classifiers.

Anuranjeeta et al. [20] performed segmentation of breast histopathological images and the segmented region's features were extracted. Based on the features extracted the images are classified as cancerous and non-cancerous using various algorithms viz J-Rip, SMO, MLP, LMT, Random forest, Rotation forest, Naïve Bayes, PART and the highest accuracy of 85.7% was obtained for Rotation forest

B. DEEP LEARNING APPROACHES

Veta et al. [45] performed an assessment to comprehend the extensive use of breast histopathology images for detecting breast cancer. Even with improvements in tissue preparation, staining, and slide digitization, inherent variations still exist, requiring the creation of strong analysis approaches to address these fluctuations. Qi et al. [46] proposed a deep active learning framework for the binary classification of breast histopathological images using confidence boosting and entropy-based techniques. Confidence boosting enhanced the selection of high confidence samples than entropy resulting in better classification. Liu et al. [47] proposed a collaborative transfer network (CTransNet) that includes transfer learning backbone for extracting

and predicting from the optimized fused features attaining an accuracy of 98.29% for multi-class classification. Xu et al. [48] proposed a Stacked Sparse Autoencoder to determine distinguishing features of nuclei from pixel intensities of histopathological images. Sliding window operation is performed over images to represent image patches using high-level features derived from the auto-encoder to effectively classify as either nuclear or non-nuclear. Chowdhury et al. [21] analysed breast histopathology images using transfer learning based on ResNet10, which had 99.58% accuracy rate in cancer early detection. Zhang et al. [22] combined CNN with graph convolutional network (GNN) where eight convolution layer was set up initially. The model improved by adding batch normalization and dropout where instead of max pooling stochastic pooling was used. Both image level features and relation awareness features were extracted which improved the accuracy of the model. The dataset used was of homogeneous smaller size, so further studies involving larger datasets is necessary to substantiate the results. For the diagnosis of breast tumors, a CNN based on residual learning with feature learning at different levels of abstraction was proposed by Gour et al. [23]. 84.34% accuracy was obtained for the binary classification and the accuracy improved to 92.52% with augmented data. Alzubaidi et al. [24] performed four class classifications of biopsy images using transfer learning in which the training is done initially on a dataset from the same domain and then on the target dataset. The reverse approach in which training on a different domain dataset and then on the target dataset is also perfomed. This method achieved an image classification accuracy of 97.4%. Ting et al. [25] used CNN for two class classification of breast mammography images and an accuracy of 90.50% was attained.

Bardou et al. [26] designed a CNN for the breast histology image classification. Accuracy for the binary classification ranged between 96.15% and 98.33% and accuracy for the multiclass classification ranged between 83.31% and 88.23%. Using deep CNN, Mahmoud [27] attained an accuracy of 86.8% for binary classification. Alom et al. [28] proposed the Inception Recurrent Residual model and obtained a testing accuracy of 99.05% for binary classification and 98.59% for multiclass classification. CNN proposed by Gecer et al. [29] obtained an accuracy of 55% for the multiclass classification of images.

Araùjo et al. [30] analyzed both binary classification and multiclass classification with an accuracy of 77.8% for four class classification and an accuracy of 83.3% for two class classification. Single-task and multi-task CNN for magnification-independent breast histopathological images were proposed by Bayramoglu et al. [31]. Single-task model achieved a recognition rate of 83.25% whereas multi-task model achieved a recognition rate of 82.13%. A structured deep-learning model for multiclass classification was proposed by Han et al. [32], which attained an average accuracy of 93.2%. Li et al. [33] performed four class classification with an accuracy of 95%.

Aljuaid et al. [34] performed binary and multiclass classification of breast histopathological images using ResNet, InceptionV3Net, ShuffleNet and transfer learning approach. Maximum average accuracies between 97.81% and 99.70% were obtained for ResNet, for binary and multiclass classification. Majumdar et al. [35] proposed an ensemble method using gamma function by combining the output of GoogleNet, Mo-bileNetV3_Small and VGG11. BreaKHis dataset has been used for evaluation and obtained results of 96.16%, 98.67%, 98.24% and 99.16% for 400X, 200X, 100X, and 40X levels of magnification and an accuracy of 96.95% on ICIAR-2018 dataset. Tayel et al. [36] used CNN techniques for detecting and localizing breast tumors.

Using transfer learning on a modified DarkNet-53 and optimized reformed differential evaluation (RDE) and reformed gray wolf (RGW) algorithms, Jabeen et al. [37] classified breast cancer on ultrasound images with a 99.1% accuracy rate. A hybrid deep learning model for feature extraction and binary classification of breast histopathology images was proposed by Wang et al. [38]. This method obtained 86% F1 score, 85% precision, 86% accuracy and 85% sensitivity. Sahu et al. [39] proposed hybrid convolutional neural networks based on ResNet18, AlexNet, ShuffleNet, VGG16, MobileNetV2 and both mammogram and ultrasound datasets were used for validation. The model gave better performance even for small datasets.

Altameem et al. [40] proposed an ensemble approach in which fuzzy rankings of the base classification techniques like DenseNet121, VGG-11, ResNet-164 and Inception V4 are built using Gompertz function and the final predictions are constructed by adaptively combining the decision scores of the base models. Gonçalves et al. [41] used particle swarm optimization and genetic algorithm for finding good hyperparameters and architectures for the DenseNet-201, ResNet-50 and VGG-16. All the three models attained F1-score above 0.90. Singh and Kumar [42] proposed a model based on inception and residual networks for the binary classification of breast histopathological images and obtained a training accuracy of 0.9642. Abunasser et al. [43] proposed a model for multiclass classification using Xception. Generative Adversarial networks (GAN) are used to increase the size of the dataset. The model obtained an F-score of 97.58%. Table 1 provides a comparison of the various deep learning classification approaches found in the literature.

III. PROPOSED METHODOLOGY
A Levenberg Marquardt based custom convolutional network (LMHistNet) was proposed for the binary as well as multiclass classification of breast cancer histopathological images. The proposed model comprises of Convolved Group Normalised block (CGN_B), dual Spatial Factorisation and Grid Reduction (SFG_{R1}&SFG_{R2}) Blocks, Convolution Block Attention (CBA_M) Module, and BN block. Optimization is performed using Levenberg Marquardt optimization. Convergence time has been significantly reduced by normalizing the

input features on models using batch normalization. Training is improved by reducing the internal covariance shift. Hinge loss function is used for better convergence. Magnification-dependent and magnification-independent classifications are efficiently performed by extracting diversified features from histopathological images.

A. LMHistNet ARCHITECTURE
This section describes how the Convolution Block Attention (CBA_M) Module is integrated with Convolved Group Normalised block (CGN_B) and dual Spatial Factorisation and Grid Reduction (SFG_{R1}&SFG_{R2}) Blocks. Convolved Group Normalised block (CGN_B) divides the channels into groups and normalized the features within groups. (SFG_{R1}&SFG_{R2}) blocks performs factorization and grid reduction thereby enhancing convergence and reduces computational cost. LMHistNet incorporates Convolution Block Attention module (CBA_M) for adaptive feature refining. The overall workflow of the proposed system is shown in figure 1.

B. CONVOLVED GROUP NORMALIZED BLOCK (CGN_B)
The input histopathological images are fed to the Convolved Group Normalised (CGN_B) block enhanced with group normalization and ReLU. The position-invariant local features generated using convolution operation is grouped and normalized by the group normalisation. The weighted features of local regions and filters of input image g and filter k resulting in convolved features as shown in equation 1 where x and y refers to the rows and columns of the features.

$$C_l = (g * k)\left[x, y\right] = \sum_j \sum_k k\left[j, k\right] g[x - j, y - k] \quad (1)$$

The extraction of features is performed using a filter size of 5×5 resulting in a recognizing pattern at various scales to create larger perceptions of objects, resulting in multiscale feature learning. The extracted convolved features $G_l[x, y]$ are fed to an average pooling layer which eliminated the spatial information, thereby enhancing its resilience to spatial translations, hence mitigating the issue of overfitting. Global average pooling obtained a reduced spatial feature G_{ap} for a convolved feature map C_l as shown in equation 2 where a, b, c is the feature size, s_l refers to the sliding stride of size 2.

$$G_{ap} = (a - k + 1)/s_l \times (b - k + 1)/s_l \times c \quad (2)$$

The resultant reduced convolved feature G_{ap} is fed to the group normalization layer G_N that allows distinct distribution to be learned for each group of channels. Group normalised features G_N are computed using mean and variance as shown in equation 3 and 4, where ε is a small constant, s_i is the set of pixels for which the mean and standard deviation are evaluated for n, the size of the set

$$Mean\ (\mu_i) = \frac{1}{n} \sum_{k \in s_i} y_k \quad (3)$$

$$Standard\ deviation\ (\sigma_i) = \sqrt{\frac{1}{n} \sum_{k \in s_i} (y_{k-\mu_i}^2 + \varepsilon} \quad (4)$$

TABLE 1. Comparison of various deep learning approaches.

Ref	Data Set	Image Count	Image Size	Method	Advantage and Disadvantage	Accuracy (%)	Computational Resources
[21]	Public dataset of Histopathological images	198,738 positive and 78,786 negative	512 to 1024	Transfer learning	Enhanced performance even with unbalanced data	99.58%	Google Colab
[24]	BACH 2018	400	2040 X 1536.	Hybrid deep CNN	Improved feature representation with huge computational complexity	97.4%	Matlab2018 RAM - 32 GB and the GPU -8 GB.
[25]	MIAS	21 benign, 17 malignant and 183 normal	1024 x 1024	CNN	Interactive detection even with small dataset	90.5%	Not mentioned
[27]	Break His	2,480 benign 5,429 malignant	700X460	Deep CNN	Integrated deep learning with pre trained models to enhance learning of complex patterns	86.8%	Not mentioned
[31]	Break His	2,480 benign 5,429 malignant	700X460	CNN	Independent from microscopy magnifications that requires single training	83.02%	Not mentioned
[34]	Break His	2,480 benign 5,429 malignant	700X460	Transfer learning	Diversified approach for multiple magnification factors.	97.81% to 99.70%	Not mentioned
[37]	BUSI	780	500 × 500	DarkNet-53	Optimized feature selection	99.1%	MATLAB2020b Intel Core i7 with 8GB 16GB RAM.
[38]	WSI and Kaggle	78,786 IDC and 198,738 non-IDC	2040 × 1536	CNN-GRU model	Hybrid deep learning model integrated with GRU to address memory limitations.	86.21%	Intel Core i7 CPU and an NVIDIA graphics processing unit (GPU).
[42]	BHI & BreakHis	2,480 benign and 5,429 malignant	700X460 pixel	Hybrid Deep Learning Model	Advance multilevel feature map for histopathology images	80.17%(BHI) 96.42% (BreakHis)	Not mentioned
[43]	Break His-	2,480 benign 5,429 malignant	700X460	CNN	Better generalization with diverge set of images generated using GAN	98.28%	Not mentioned

Convolved group normalization block (CGN_B) incorporates convolutional layer augmented by the group normalization subsequently followed by ReLu activation enhances the network stability to facilitate the learning of intricate spatial information features. Local patterns like edges and textures effectively capture hierarchical and spatial patterns that differentiate benign from malignant features. Complex patterns are efficiently generated through non-linearity using activation functions and utilizing group normalization. The output of the group normalization G_N is optimized using ReLU activation function and the resulting normalized feature CGN_O is subjected to the Spatial Factorisation and Grid Reduction (SFG_R) Block

C. SPATIAL FACTORISATION AND GRID REDUCTION (SFG_{R1}) BLOCK

The LMHistNet includes two identical spatial factorisation and grid reduction (SFG_{R1}, SFG_{R2}) blocks as shown in Figure 1. The Spatial Factorisation and Grid Reduction ((SFG_{R1}) block is shown in figure 2. The CGN_B layer output (CGN_O) characteristics are fed into the SFG_{R1} block. The convolved features are factorised into smaller size through the advantage of auxiliary units leading to wider network, overcoming vanishing gradient problem. Multiple filters of varying sizes on the same level, in parallel units are stacked rather than deep layers, has the potential to decrease the dimensions of the network and expedite computations.
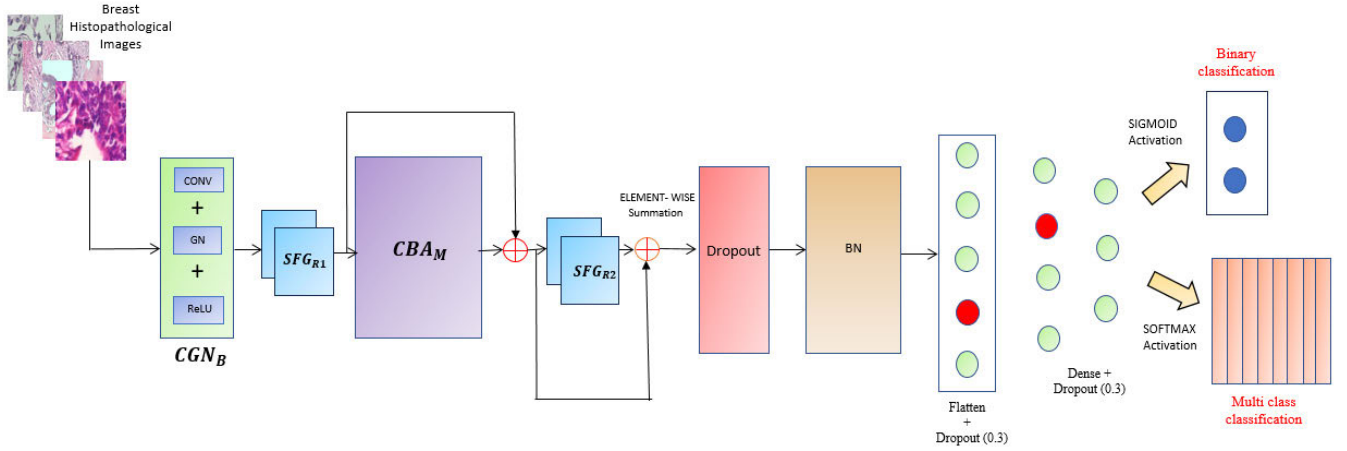
FIGURE 1. Overall workflow of the proposed system.

The process of breaking down a convolution into smaller convolutions resulted spatial factorization, involving the use of asymmetric convolutions attaining optimising grid size reduction.
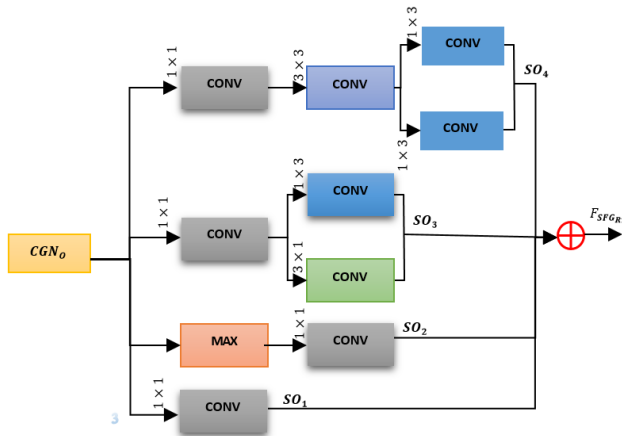

FIGURE 2. Spatial factorisation and grid reduction (SFG_{R1}) block.

The normalized feature output CGN_O from the Convolved Group Normalised (CGN_B) block is processed in a stacked manner involving four stack operations $SO_{i=1to4}$. SO_1 involves a convolution layer with a filter size of 1×1 that enables dimensionality reduction with non-linear feature generation that supports squeezing to a desired output depth (equation 5). The output obtained from Convolved Group Normalised Block (CGN_O) is subjected to convolutional layer ($_L$) with 1×1 filter resulting in reduced SO_1 features by primarily focusing on depth dimension rather than spatial features.

$$SO_1 = (C_L (CGN_O))^{1 \times 1} \qquad (5)$$

Similarly, the output obtained from Convolved Group Normalised Block (CGN_O) is subjected to maxpooling followed by convolutional layer with 1×1 filter resulting

in reduced SO_2 features as shown in equation 6. Applying max pooling (P_M), followed by a convolution layer (C_L) with a filter size of 1×1, decreases the spatial dimensions of the feature maps by improving the non-linear interactions between features resulting complex patterns.

$$SO_2 = (P_M(C_L (CGN_O)))^{1 \times 1} \qquad (6)$$

SO_3 maps the features through a 1×1 convolved layer (C_L) and the resultant feature is processed using dual filters of size 3×1 and 1×3 resulting in spatial factorization into asymmetric convolutions as 3×3 is split into 1×3 and 3×1 with reduced complexity(equation 7).

$$SO_3 = (C_L(C_L (CGN_O))^{1 \times 1})^{1 \times 3} + (C_L(C_L (CGN_O))^{1 \times 1})^{3 \times 1}) \qquad (7)$$

SO_4 processes the CGN_O using an auxillary convolution layer of 1×1 followed by 3×3 filter further processed by 3×1 and 1×3 factored spatial feature map (equation 8) resulting in spatial structural pattern précised towards specific orientation.

$$SO_4 = (C_L(C_L(C_L (CGN_O))^{1 \times 1})^{3 \times 3})^{1 \times 3} + (C_L(C_L(C_L (CGN_O))^{1 \times 1})^{3 \times 3})^{3 \times 1} \qquad (8)$$

The depth features obtained from SO_1, complex patterns of SO_2, SO_3 and SO_4 are concatenated resulting in expressive diverse feature $F_{SFG_{R1}}$ (equation 9) balancing weak contributions from shallow layers.

$$F_{SFG_{R1}} = SO_1 + SO_2 + SO_3 + SO_4 \qquad (9)$$

Spatial factorization and grid reduction blocks (SFG_{R1} & SFG_{R1}) involves standard convolution, pooling and depth-wise separable convolutional layers that reduces network dimensions leading to a computationally efficient network through feature refinement which helps overcoming vanishing gradient problem. SFG_{R1} & SFG_{R1} achieves translation invariance features by downsampling the spatial dimensions

through the separation of spatial and depth-wise convolutions enhancing robust generalization by capturing both local and global features thereby contributing to model stability.
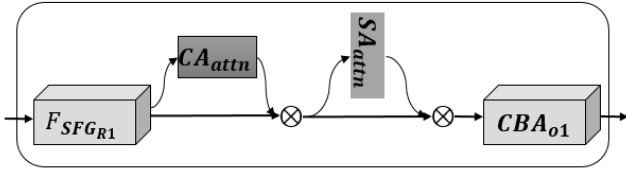

FIGURE 3. Convolution block attention (CBA_M).

D. CONVOLUTION BLOCK ATTENTION (CBA_M) MODULE

The figure 3 shows the Convolution Block Attention (CBA_M) which involves channel attention (CA_{attn}) module and spatial attention module (SA_{attn})), the result of with which the network can concentrate more on discriminative features.

1) CHANNEL ATTENTION (CA_{attn}) MODULE

The feature map $F_{SFG_{R1}}$ from Spatial Factorisation and Grid Reduction (SFG_R) Block is fed to Channel Block Attention (CB_{attn}) Block which exploits the inter-channel relationship of features as shown in figure 4.
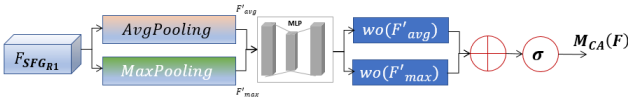
FIGURE 4. Channel attention (CA_{attn}) module.

Channel attention involves each channel of a feature map as a feature detector, to emphasize the significance of the information present within the input feature. To compute the channel attention efficiently, the input feature map's spatial dimension is compressed. The spatial information of a feature map is initially aggregated through the utilization of average-pooling and max-pooling processes resulting in average F'_{avg} and max pooled features F'_{max}. The F'_{avg} and F'_{max} are fed to Multilayer perceptron (MLP) network and the resulting channel attention map is one dimensional represented as $M_{CA} \in \mathbb{R}^{C \times 1 \times 1}$ where c refers to the number of channels.

The multilayer perceptron with one hidden layer is applied to both the descriptors and the element wise summation is used to merge the output feature vectors. To reduce parameter overhead, the hidden activation size is set to $\mathbb{R}^{C/r \times 1 \times 1}$ where r is the reduction ratio. The channel attention is computed as shown in equation (10), Where σ is a logistic function ranging between 0 and 1, irrespective of the input given and $W_0 \in \mathbb{R}^{C/r \times C}$ and $W_1 \in \mathbb{R}^{C \times C/r}$

$$M_{CA}(F) = \sigma(W_1(W_0(F'_{avg}) + W_1(W_0(F'_{max}))) \quad (10)$$

2) SPATIAL BLOCK ATTENTION (SA_{attn}) BLOCK

Channel information of a feature map $M_{CA}(F)$ are fed to the Spatial Attention (SB_{attn}) Block where max pooling and average pooling operations along the channel axis are aggregated, thereby generating, average pooled features as $F'_{avg} \epsilon \mathbb{R}^{1 \times H \times W}$ and max pooled features $F'_{max} \epsilon \mathbb{R}^{1 \times H \times W}$ that are concatenated and convolved by a convolution layer with filter size 7×7 to produce the two dimensional spatial attention map. The spatial map is denoted as equation (11) where $f^{7 \times 7}$ indicates convolution operation filter size 7×7.

$$M_{SA}(F) = \sigma(f^{7 \times 7}([F'_{avg}; F'_{max}])) \quad (11)$$

$$CBA_{o1} = M_{CA}(F) + M_{SA}(F) \quad (12)$$

The obtained CBA_{o1}(equation12) features which is the output of (CBA_M) Module is further processed with the factorised feature $F_{SFG_{R1}}$. inorder to avoid losing any important characters. The feature map $F_{SFG_{R1}}$., and the output obtained by element wise multiplication of CBA_{o1} and $F_{SFG_{R1}}$., are added to generate (CBA_O) resulting in improved robustness and interpretability (equation 13).

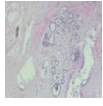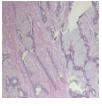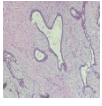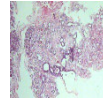$$CBA_0 = F_{SFG_{R1}} + F_{SFG_{R1}} \otimes CBA_{o1} \quad (13)$$
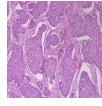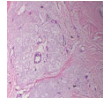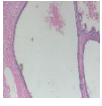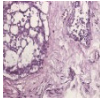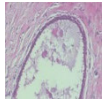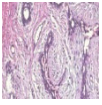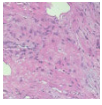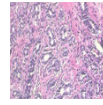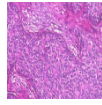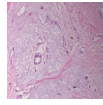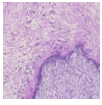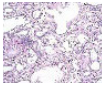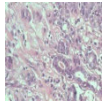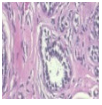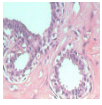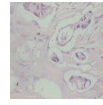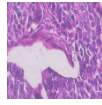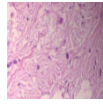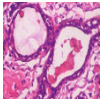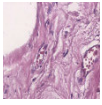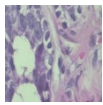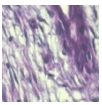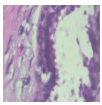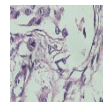
The channel and spatial features are generated attaining an enhanced specific pattern of features to effectively determine the benign and malignant characteristics in the histopathological data. A channel block attention module (CBA_M) facilitates the integration of channel and spatial information to generate meaningful and informative feature representations surpassing the impact of redundant or less informative features. This attention module focuses on specific channels to retrieve more discriminative information by dynamically analyzing the characteristics of the input data thus enabling the model to learn and effectively exploit the complex relationships within the data. The feature map $F_{SFG_{R1}}$ along with channel attention map CBA_{o1} is fed to the $SFG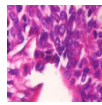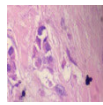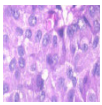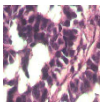_{R2}$ block. Dropout is applied to the summed up output of SFG_{R2} and attention map CBA_0 with a probability of 0.3 which prevented the co adaptation of neurons. Batch normalization is applied which reduced the change in the distributions of networks activations during training.

E. OUTPUT LAYERS

The batch normalized attention map is fed to flatten layer which converted the output features into a single long continuous vector. Dropout was implemented in the flatten layer by selecting nodes to dropout with a probability of 30% which improved the model generalization and reduced overfitting. The flattened output is processed to a dense layer with a dropout of 30% and ReLU activation function is used here. For binary classification, the activation function used is sigmoid in the outer dense layer and for multiclass classification, softmax activation function is used as given in equation 14, where, σ is the softmax, $\vec{z}$ is Input vector, e^{z_l} is the standard exponentiation function which refers to vector of each input element and e^{z_m} refers to the vectors normalized output for 'm' number of classes.

$$\sigma(\vec{z})_i = \frac{e^{z_l}}{\sum_{j=1}^{n} e^{z_m}} \quad (14)$$

TABLE 2. Images of eight different classes with four different resolutions from BreaKHis dataset.

	Adenosis	Fibroadenoma	Phyllodes Tumor	Tubular adenoma	Ductal carcinoma	Lobular carcinoma	Mucinous carcinoma	Papillary carcinoma
40x								
100x								
200x								
400x								

F. LEVENBERG MARQUARDT OPTIMIZATION AND HINGE LOSS FUNCTION

The proposed model is enhanced using Levenberg Marquardt learning the where the error is computed by comparing the labeled data and the weight updation enhanced by the combined Error Backpropagation and Gauss Newton algorithm [44]. The Error Backpropagation algorithm is inefficient due to slow convergence which is improved using the Gauss Newton algorithm though divergent approximation. The Levenberg Marquardt algorithm perform well in terms of speed and stability and always converges well albeit the error surface is far more complex. The weight updation of steepest descent algorithm, gauss newton algorithm and Levenberg Marquardt is given as equation 15, 16 and 17.

$$w_{i+1} = w_i - \lambda g_i \qquad (15)$$

$$w_{i+1} = w_i - (J_i^T J_i)^{-1} J_i e_i \qquad (16)$$

$$w_{i+1} = w_i - (J_i^T J_i + \lambda I)^{-1} J_i e_i \qquad (17)$$

where w_i, w_{i+1} refers the previous and current weight, λ refers to combination coefficient, 'I' refers identity matrix and J refers to Jacobin matrix. g is the first order derivative of the error function and e is the error vector of the ith iteration. Combination coefficient λ gets reduced if parameter yields fewer errors than the previous one based on weight updation and the resulted error. The Levenberg–Marquardt optimization method is used to improve the speed of model training and facilitate efficient convergence of the network by adjusting step size and damping adaptively. Hinge loss

function used in the model improved the training speed of the model and also resulted in better convergence due to its linear property (equation 18) where hl is the hinge loss, a_v is the actual value and p_v is the predicted value.

$$hl = \max(0, 1 - a_v.p_v) \qquad (18)$$

Hinge loss for multiclass classification is defined as equation 19, where $w_{pv}x$ and $w_{av}x$ are model parameters.

$$hl = \max(0, 1 + (w_{pv}x - w_{av}x)) \qquad (19)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section details the experimental setup, dataset description, augmentation methods, ablation study, and performance evaluation of the breast cancer using histopathological images classification using various performance metrics.

A. DATASET DESCRIPTION

A publicly available dataset named BreaKHis is used in the proposed work which contains microscopic images of breast tissues collected from 82 patients. The dataset includes 2,480 benign samples and 5429 malignant samples in four different magnifying factors- 40X, 100X, 200X and 400X. Malignant tumors are cancerous whereas benign tumors are noncancerous. Table 2 shows the sample images from the BreaKHis dataset belonging to eight different classes with four different resolutions. The number of images included for eight class classification is given as follows, 444 images of Adenosis, 1014 images of Fibroadenoma, 453 images of Phyllodes tumor, 569 images of Tubular

adenoma, 3457 images of Ductual carcinoma, 626 images of Lobular carcinoma, 792 images of Mucinous carcinoma, and 580 images of Papillary carcinoma. 80% data of the dataset is used for training and 20% for testing.

B. EXPERIMENTAL SETUP

The proposed task was executed on a cloud computing environment called Google Colab notebooks. Google Colab provides a tensor processing unit (TPU) and graphics processing unit (GPU) for creating a deep neural learning model. This cloud-based system made it easier for the deep learning model to be trained and run effectively. LMHistNet model was implemented using Python programming and several essential libraries were utilized, including Keras, matplotlib, Tensor flow, OS, and sklearn, during the implementation. The configuration of the system used to experiment the proposed work is shown in Table 3.

TABLE 3. Computer configuration of the proposed LMHistNet.

Item	Configuration
Processor	Intel (R) Xeon (R) Gold 6230 CPU @2.10GHz
Graphics Card	NVIDIA Quadro RTX 5000 16 GB
Ram Size	64 GB
Hard-Disk Size	2 TB

C. DATA PREPROCESSING AND AUGMENTATIONS

The input image size of the network structure is resized to 224 X 224. The LMHistNet model's convergence and the generation of standard features will be significantly impacted by the images' varying size. Data augmentation is utilized to artificially enlarge the dataset's size by performing various operations like translation, rotation, shear, etc. But in the case of our images all the data augmentation techniques would not work as expected when compared to natural images. Hence prudent decisions must be taken while selecting the data augmentation techniques. In our experiment images are rotated with a rotation range equal to 15 and the pixels move outside the image are filled by fill_mode to nearest which replaces the empty area with the neighboring pixel. The images after data augmentation are shown in figure 5.

D. HYPERPARAMETER TUNING

The learning rate, dropout rate, batch size, epochs and optimization units employed in the gradient were among the hyperparameters on LMHistNet models that were tuned using grid search. Based on experimentation, a batch size as 32, and learning rate as 0.001 were used in the proposed work LMHistNet. Levenberg–Marquardt optimizer is used for fine-tuning network weights for each of the deep neural networks. ReLU activation function was used in the intermediate layers
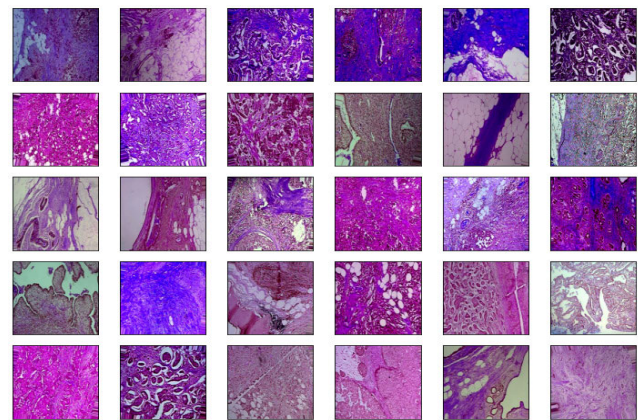

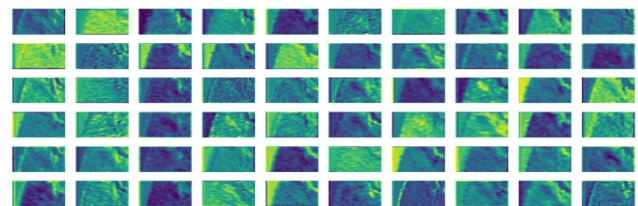**FIGURE 5.** Sample data augmented images.

FIGURE 6. Visual interpretation of the intermediate features.

TABLE 4. Classification accuracy of LMHistNet.

Proposed Model	Accuracy	
	Binary	Multiclass
CGN_B	95%	85.55%
$CGN_B + SFG_{R1}$	95.53%	86.33%
$CGN_B + SFG_{R1} + CBA_M + SFG_{R1}$	97.65%	87.65%
Proposed LMHistNet	99%	88%

and the final dense layer was used with the Softmax function for output results. The model was trained for 200 epochs. 30% dropout helped to minimize the overfitting to a greater extend achieving a high accuracy for multiclass classification.

E. VISUAL INTERPRETATION OF LMHistNet MODEL FEATURES

Figure 6 displays the features retrieved by the LMHistNet inner layers during the training phase. The visual representation of the intermediate feature layers shows how the extracted features are used to train the model. This depiction greatly improves our comprehension of how the LMHistNet model interprets the breast cancer histopathologial images internally.

F. EVALUATION METRICS OF THE PROPOSED ARCHITECTURE

The proposed LMHistNet model is evaluated using various parameters like accuracy, precision, recall and f-score. Positive and negative are two terms used for classification

TABLE 5. Model accuracy and loss obtained for the binary classification of magnification-dependent and independent classes.
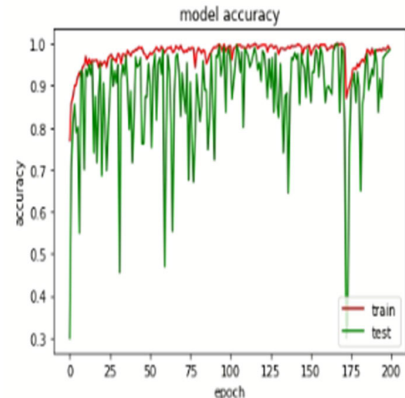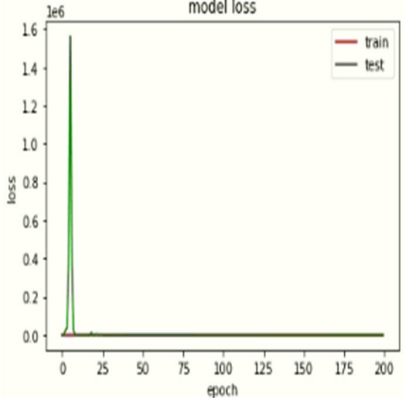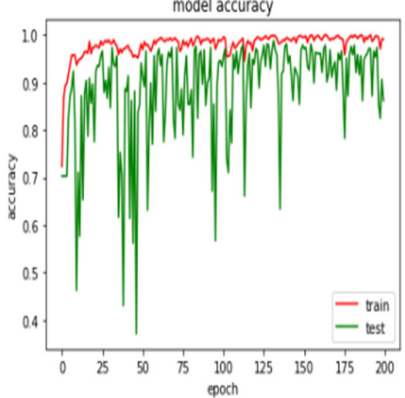
Image Resolution	Accuracy	Loss
Magnification Independent	model accuracy	model loss
40X	model accuracy	model loss
100X	model accuracy	model loss
200X	model accuracy	model loss
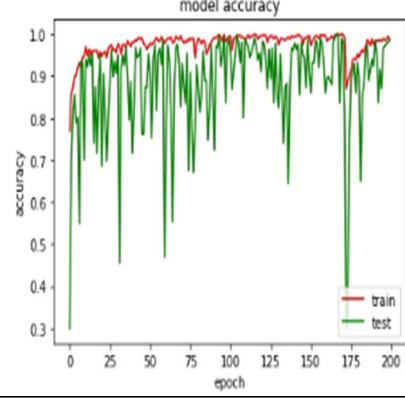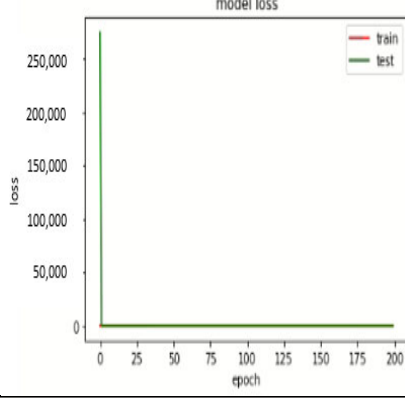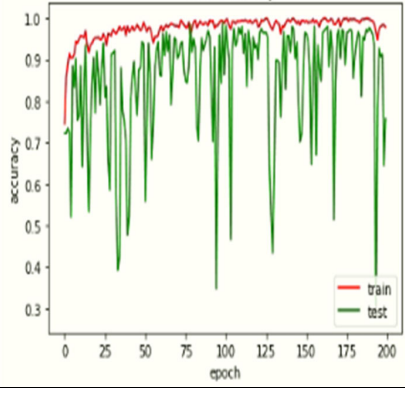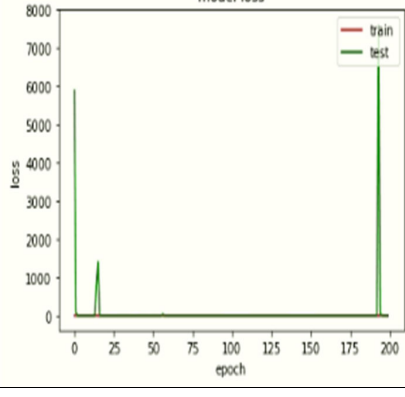
TABLE 5. *(Continued.)* **Model accuracy and loss obtained for the binary classification of magnification-dependent and independent classes.**
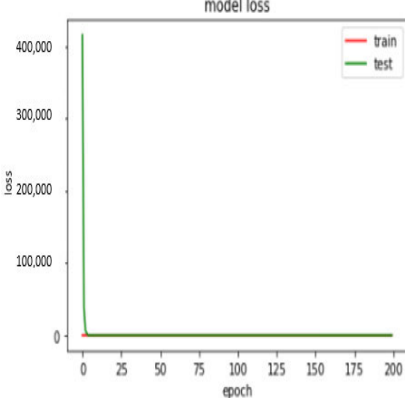

TABLE 6. **Image count of four different resolution.**

	Adenosis	Fibro Adenoma	Phyllodes Tumor	Tubular Adenoma	Ductual Carcinoma	Lobular Carcinoma	Mucinous Carcinoma	Papillary Carcinoma
40X	114	253	109	149	870	156	205	162
100X	113	260	121	150	903	170	222	142
200X	111	264	108	140	896	163	196	135
400X	106	237	115	130	788	137	169	141

purpose and positive in our experiment means the image is malignant or cancerous. The equation for each is given below. Accuracy is calculated using equation (20) and f1-score using equation (21). Precision for binary classification is found using equation (22) and for multiclass classification using equation (23). TP is the number of correctly diagnosed breast cancer images, TN denotes the number of correctly classified malignant images and FP denotes the number of malignant images that are classified as benign. FN stands for false negative and it refers to the number of benign images that are classified as malignant.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(20)

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall}$$
(21)

$$Precision\ for\ binary\ classification = \frac{TP}{TP + FP}$$
(22)

$$Precision\ for\ multiclass = \frac{\sum_i TP_i}{\sum_i TP_i + FP_i}$$
(23)

G. ABLATION STUDIES

Ablation studies were used to examine the effects of several LMHistNet model modules on the classification of breast histopathology images. It has been found that the performance of the model gradually improves with the addition of the blocks. The LMHistNet model was evaluated using various parameters like Accuracy, Precision, Recall, and F1-score and Confusion matrix. The proposed LMHistNet model's baseline block, CGN_B consists of convolution, group normalization, and ReLU. It achieved 95% classification accuracy for binary classification and 85.55% classification accuracy for multiclass classification. SFG_{R1} block is implemented along with CGN_B which obtained accuracies of 95.53% and 86.33% for binary and multiclass classification respectively. The model accuracy was further improved by the addition of CBA_M module which improved the discriminative power of the model by suppressing the less informative channels. Incorporating SFG_{R2} to the network rsulting in a deeper network architecture. The model is further optimized with Levenberg Marquardt resulting in 99% and 88% accuracy for binary and multiclass classification of histopathological breast cancer dataset. The results are shown in Table 4.

TABLE 7. Model accuracy and loss for multiclass classification of magnification-dependent and magnification-independent classes.
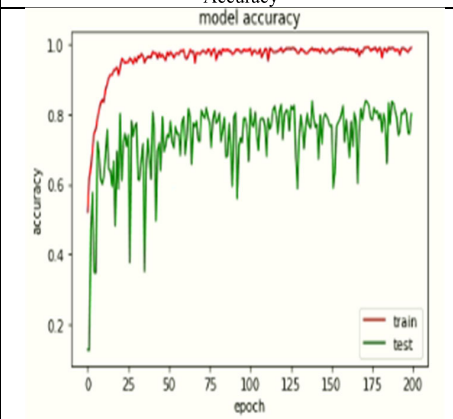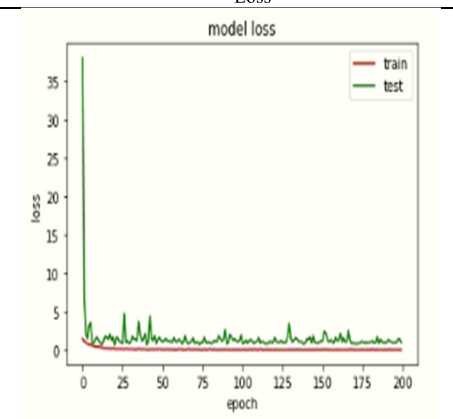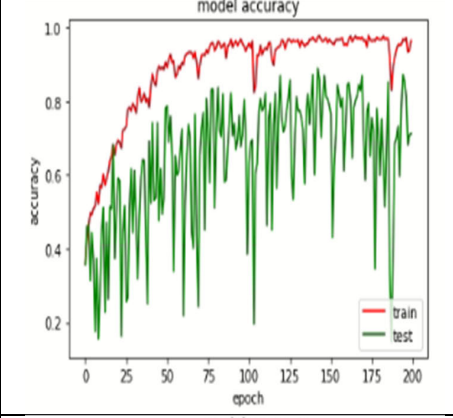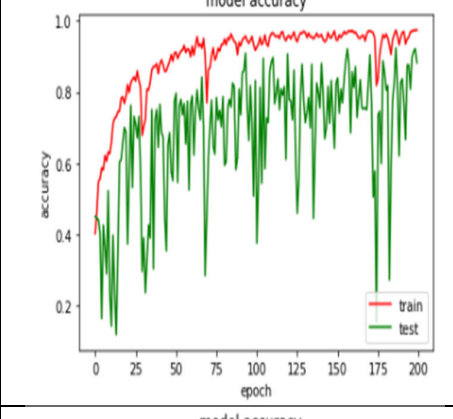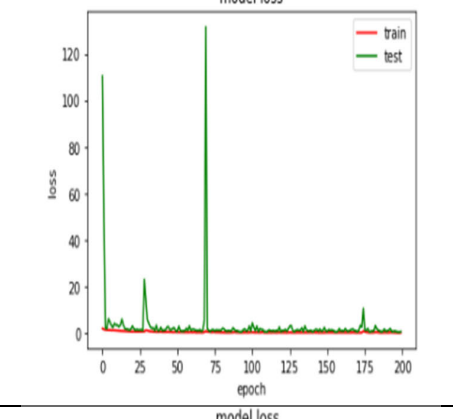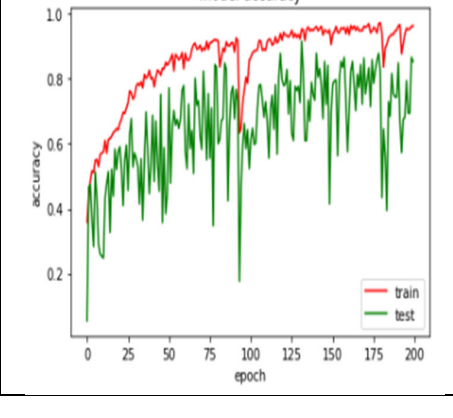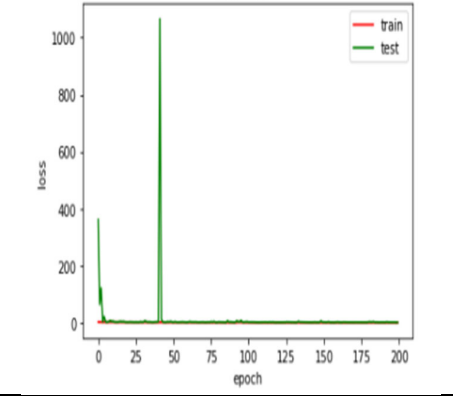
Image Resolution	Accuracy	Loss
Magnification-Independent		
40X		
100X		
200X		

TABLE 7. *(Continued.)* **Model accuracy and loss for multiclass classification of magnification-dependent and magnification-independent classes.**
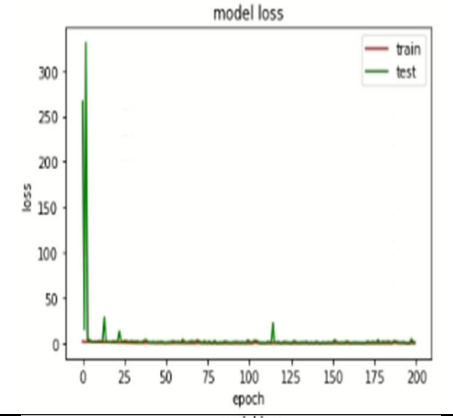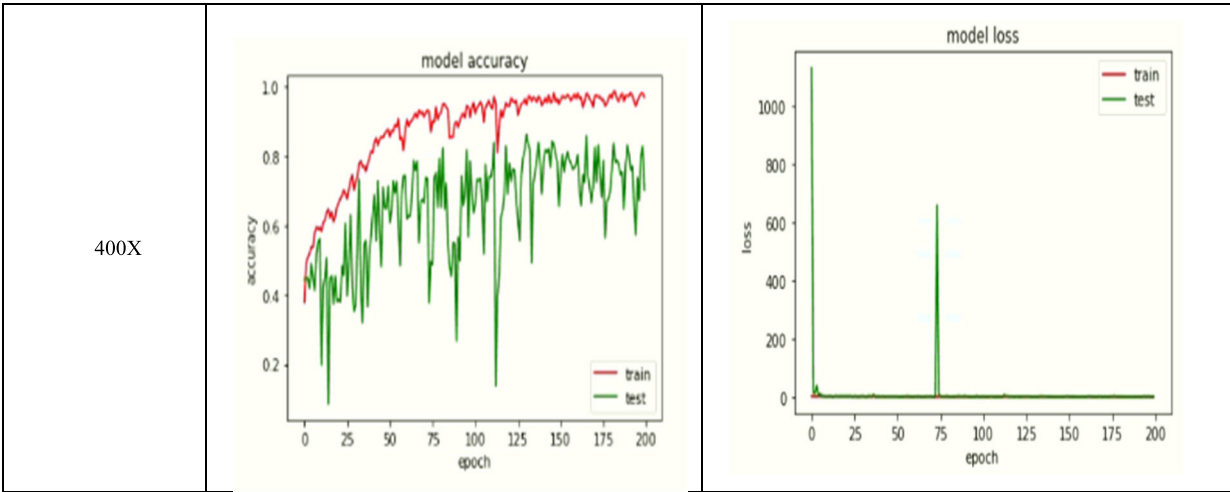

TABLE 8. **Overall performance validation of the proposed LMHistNet.**

	Resolution	Accuracy (%)	Precision	Recall	F1 score
Magnification Independent Binary Classification	All	75	77	81	74
Magnification Independent Multiclass Classification	All	79	77	72	74
Magnification Dependent Binary Classification	40X	86	88	86	85
	100X	99	99	99	99
	200X	85	82	82	81
	400X	**96**	**97**	**96**	**96**
Magnification Dependent Multiclass Classification	40X	71	81	71	73
	100X	88	89	88	88
	200X	85	86	85	85
	400X	70	80	70	71

H. MAGNIGFICATION DEPENDENT AND INDEPENDENT BINARY CLASSIFICATION

The images, irrespective of their resolutions, are classified into two classes— benign and malignant. A total of 2480 benign and 5428 malignant images were analyzed using the proposed LMHistNet model. Similarly, 625 benign and 1393 malignant of 40X resolution, 644 benign and 1437 malignant of 100X resolution, 623 benign and 1390 malignant of 200X resolution, 588 benign and 1253 malignant of 400X resolution of magnification-dependent images were also trained and tested. The Histopathological images of four different magnifications have been used viz 40x, 100x, 200x and 400x. Images of different magnifications possesses different features which creates more learning to generate good classification accuracy. Eventhough the imbalanced data is augmented and regularized to improve generalization, the use of random initializations, different dropout and batch normalization led to different convergence paths, causing fluctuations in the training process. The learning rate and the step size was fine tuned to reduce fluctuations to minimal obtaining better accuracy for all magnifications. The model accuracy and loss obtained for the binary classification of both magnification-dependent and independent classes are shown in Table 5.

I. MAGNIGFICATION DEPENDENT AND INDEPENDENT MULTICLASS CLASSIFICATION

The magnification-dependent and independent classifications into eight classes were analyzed using the proposed LMHistNet model. The eight different classes and the number of images trained and tested are as follows: 444 images of adenosis, 1014 images of fibroadenoma, 453 images of phyllodes tumor, 569 images of tubular adenoma, 792 images of mucinous carcinoma, 626 images of lobular carcinoma, 3457 images of ductal carcinoma and 580 images of papillary carcinoma The number of images belonging to eight different classes with four different resolutions are given in Table 6. The model accuracy and loss obtained for magnification dependent and independent multiclass classification is shown in table 7.

J. PERFORMANCE ANALYSIS

The LMHistNet model for breast histopathology image classification attained 75% magnification-independent binary classification accuracy and 79% magnification-independent multiclass image classification accuracy. The proposed model achieved a precision of 77 and an F1 score of 74 for the magnification-independent binary and multiclass classifications, whereas the recall value for the binary

classification is 81, which is better than the recall value for the magnification-independent multiclass classification.

LMHistNet attained a 99% accuracy for the magnification-dependent binary classification for histopathological images with a resolution of 100X, which is higher than the 86%, 85% and 96% accuracies achieved for images with resolutions of 40X, 200X and 400X, respectively. The histopathological images with a resolution of 100X had the highest precision, recall and F1 score values of 99. The LMHistNet model achieved a precision of 88, recall of 86 and F1 score of 85 for the magnification-dependent binary classification of images with a resolution of 40X, while the precision, recall and F1 score were 82, 82 and 81 for the histopathological images with a resolution of 200X. For the magnification-dependent binary classification of histopathological images with a resolution of 400X, the LMHistNet model achieved a precision, recall and F1 score of 97, 96 and 96 respectively.

With an accuracy of 88% and precision, recall and F1 score, of 81, 71 and 73 respectively, the histopathological images at a resolution of 100X received the highest classification accuracy for magnification-dependent multiclass classification using LMHistNet. In the magnification-dependent multiclass classification, images with a resolution of 200X had the second-highest accuracy of 85%, followed by images with resolutions of 40X and 400X, which had accuracies of 71% and 70%, respectively. For the magnification-dependent multiclass classification of images with a resolution of 40X, the LMHistNet model achieved an accuracy of 81%, recall of 71, F1 score of 73 and values of 85, 86 and 85 respectively, for the precision, recall and F1 score of images with a resolution of 200X. The model achieved the precision, recall and F1 score values of 80, 70 and 71 respectively, for histopathological images at a resolution of 400X. A summary of the results obtained for the binary as well as multiclass classification of images with resolutions of 40X, 100X, 200X and 400X are shown in Table 8.

K. COMPARISON OF THE PROPOSED LMHistNet WITH STATE-OF-THE-ART ARCHITECUTRES

Chowdhury et al. [21] performed binary histopathological image classification using the BreaKHis dataset with an accuracy of 99.58%. Gour et al. [23] achieved a binary classification accuracy of 92.5% and used the same dataset. Alzubaidi et al. [24] attained a four class classification accuracy of 97.4%. Mahmoud [27] and Anuranjeeta et al. [20] also used the same BreaKHis dataset for histopathological image analysis and the accuracies obtained were 86.8% and 85.7% respectively. Alom et al. [28] and Bardou et al. [26] performed both binary as well as multiclass classifications, with accuracies 98.33% and 99.05 for the binary classification and 88.23% and 98.59% for the multiclass classification respectively. Gecer et al. [29] performed five class classification of histopathological images and obtained an accuracy of 55%. Araùjo et al. [30] performed both binary as well as four class classification and acquired the results of 83.3% and 77.8% for binary and four

TABLE 9. Comparison of LMHistNet with state of the art architectures.

Reference	Accuracy	
	Binary classification	Multiclass classification
[21]	99.58	-
[23]	92.5	-
[26]	98.33	88.23%(eight class)
[24]		97.4%
[27]	83.33	-
[29]	-	55%(five class)
[30]	83.3	77.8(four class)
[31]	83.25	-
[32]	-	93.2%(eight class)
[33]	-	88.89%(four class)
[34]	99.70%	97.81%
[35]	99.16	-
[37]	-	99.1(three class)
[38]	86.21%	-
[39]	-	99.17%(three class)
[40]	-	99.32%(three class)
[42]	0.9642	-
[43]	-	97.6%(EIGHT class)
LMHistNet	99%	88%

class classification, respectively. Bayramoglu et al. [31] performed binary classification with an accuracy of 83.25% and Han et al. [32] performed eight class classification of breast histopathology images with an accuracy of 93.2%. Li et al. [33] attained an accuracy of 88.89 percentage in the four-class image classification. Aljuaid et al. [34] performed binary and multiclass classification of breast histopathology images using ResNet 18, ShuffleNet, and Inception-V3Net and obtained best results for ResNet. Majumdar et al. [35] perfomed binary classification of breast hisopathological images using rank based ensembling and obtained maximum accuracy of 99.16%. Jabeen et al. [37] performed classification of breast ultrasound images with a maximum accuracy of 99.1%. Wang et al. [38] performed binary classification of Whole slide images with precision 85.50%, an accuracy of 86.21%, sensitivity 85.60% and specificity 84.71 and F1-score 88%.Sahu et al. [39] performed classification of ultrasound and mammogram images into abnormal normal and benign classes with a maximum accuracy of 99.17%. Altameem et al. [40] obtained a classification accuracy of 99.32%. Singh and Kumar [42] obtained a binary classification accuracy of 0.9642. The comparison of the results is shown in Table 9.

V. CONCLUSION AND FUTURE WORK

The primary concern deals with the optimized and efficient breast histopathological image classification, thus assisting

the diagnostic process. The proposed LMHistNet model was used for binary as well as multiclass classification. Convolution block attention module is integrated with the LMHistNet model which aids the model to concentrate more on discriminative features. The Levenberg–Marquardt optimization algorithm strengthened the features by optimizing the weights effectively. Batch normalization was applied between these stacked clusters to improve the feature acquired, allowing improved categorization. The proposed LMHistNet obtained promising results for binary classification, but the results obtained for multiclass classification were comparatively less due to many factors, such as the classification into a large number of classes, i.e., eight classes; also, the number of image samples in the individual class is less. Both magnification-dependent and magnification-independent classifications have been performed using the BreaKHis dataset. The LMHistNet obtained an accuracy of up to 85% for the multiclass classification and an accuracy of 99% for binary classification.

Future work can be focused on optimizing the preprocessing techniques to enhance image quality. The availability of a huge dataset is a significant obstacle for computer-aided diagnostics. In order to move beyond the limits of tiny datasets, some shot learning models can be investigated. Deep learning models for multiclass categorization, which are similar to BRAIDS (Breast Imaging Reporting and Data System) and TNM (Tumor Node, Metastasis), can be implemented.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics," *CA Cancer J. Clinicians*, vol. 67, no. 27, pp. 7–30, 2017.

[3] K. L. Britt, J. Cuzick, and K.-A. Phillips, "Key steps for effective breast cancer prevention," *Nature Rev. Cancer*, vol. 20, no. 8, pp. 417–436, Aug. 2020.

[4] Y.-S. Sun, Z. Zhao, Z.-N. Yang, F. Xu, H.-J. Lu, Z.-Y. Zhu, W. Shi, J. Jiang, P.-P. Yao, and H.-P. Zhu, "Risk factors and preventions of breast cancer," *Int. J. Biol. Sci.*, vol. 13, no. 11, pp. 1387–1397, 2017.

[5] Centers for Disease Control and Prevention. *Content Source: Division of Cancer Prevention and Control*. Accessed: Apr. 20, 2022. [Online]. Available: https://www.cdc.gov/cancer/dcpc/about/index.htm

[6] N. Takkar, S. Kochhar, P. Garg, A. Pandey, U. Dalal, and U. Handa, "Screening methods(clinical breast examination and mammography) to detect breast cancer in women aged 40–49 years," *J. Mid-Life Health*, vol. 8, no. 1, p. 2, 2017.

[7] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," *Artif. Intell. Med.*, vol. 127, May 2022, Art. no. 102276.

[8] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network—A deep learning approach," *Proc. Comput. Sci.*, vol. 132, pp. 679–688, Jan. 2018.

[9] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[10] 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.

[11] S. Bacha and O. Taouali, "A novel machine learning approach for breast cancer diagnosis," *Measurement*, vol. 187, Jan. 2022, Art. no. 110233.

[12] B. Dai, R.-C. Chen, S.-Z. Zhu, and W.-W. Zhang, "Using random forest algorithm for breast cancer diagnosis," in *Proc. Int. Symp. Comput., Consum. Control (IS3C)*, Dec. 2018, pp. 449–452.

[13] S. Kabiraj, M. Raihan, N. Alvi, M. Afrin, L. Akter, S. A. Sohagi, and E. Podder, "Breast cancer risk prediction using XGBoost and random forest algorithm," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–4.

[14] D. A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine learning classification techniques for breast cancer diagnosis," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 495, Jun. 2019, Art. no. 012033.

[15] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement*, vol. 146, pp. 800–805, Nov. 2019.

[16] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *J. Personalized Med.*, vol. 11, no. 2, p. 61, Jan. 2021.

[17] E. Michael, H. Ma, H. Li, and S. Qi, "An optimized framework for breast cancer classification using machine learning," *BioMed Res. Int.*, vol. 2022, pp. 1–18, Feb. 2022.

[18] S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, Apr. 2021, pp. 97–101.

[19] M. A. Naji, S. E. Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine learning algorithms for breast cancer prediction and diagnosis," *Proc. Comput. Sci.*, vol. 191, pp. 487–492, Jan. 2021.

[20] A. Anuranjeeta, K. K. Shukla, A. Tiwari, and S. Sharma, "Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features," *Biomed. Pharmacol. J.*, vol. 10, no. 1, pp. 353–366, Mar. 2017.

[21] D. Chowdhury, A. Das, A. Dey, S. Sarkar, A. D. Dwivedi, R. R. Mukkamala, and L. Murmu, "ABCanDroid: A cloud integrated Android app for noninvasive early breast cancer detection using transfer learning," *Sensors*, vol. 22, no. 3, p. 832, Jan. 2022.

[22] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, "Improved breast cancer classification through combining graph convolutional network and convolutional neural network," *Inf. Process. Manage.*, vol. 58, no. 2, Mar. 2021, Art. no. 102439.

[23] M. Gour, S. Jain, and T. Sunil 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.

[24] 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.

[25] 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.

[26] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.

[27] M. Mahmoud, "Breast cancer classification in histopathological images using convolutional neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 64–68, Apr. 2018.

[28] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *J. Digit. Imag.*, vol. 32, no. 4, pp. 605–617, Aug. 2019.

[29] 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.

[30] T. Araùjo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polìonia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS One*, vol. 12, no. 6, 2017, Art. no. e0177544.

[31] 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.

[32] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.*, vol. 7, no. 1, p. 4172, Jun. 2017.

[33] 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.

[34] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Comput. Methods Programs Biomed.*, vol. 223, Aug. 2022, Art. no. 106951, doi: 10.1016/j.cmpb.2022.106951.

[35] S. Majumdar, P. Pramanik, and R. Sarkar, "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119022.

[36] M. B. Tayel, M. A. A. Mokhtar, and A. F. Kishk, "Breast cancer diagnosis using histopathology and convolution neural network CNN method," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer, 2023, pp. 585–600.

[37] K. Jabeen, M. A. Khan, M. Alhaisoni, U. Tariq, Y.-D. Zhang, A. Hamza, A. Mickus, and R. Damaševičius, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," *Sensors*, vol. 22, no. 3, p. 807, Jan. 2022, doi: 10.3390/s22030807.

[38] X. Wang, I. Ahmad, D. Javeed, S. A. Zaidi, F. M. Alotaibi, M. E. Ghoneim, Y. I. Daradkeh, J. Asghar, and E. T. Eldin, "Intelligent hybrid deep learning model for breast cancer detection," *Electronics*, vol. 11, no. 17, p. 2767, Aug. 2022, doi: 10.3390/electronics11172767.

[39] A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104292.

[40] A. Altameem, C. Mahanty, R. C. Poonia, A. K. J. Saudagar, and R. Kumar, "Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques," *Diagnostics*, vol. 12, no. 8, p. 1812, Jul. 2022.

[41] C. B. Gonçalves, J. R. Souza, and H. Fernandes, "CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105205, doi: 10.1016/j.compbiomed.2021.105205.

[42] S. Singh and R. Kumar, "Breast cancer detection from histopathology images with deep inception and residual blocks," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5849–5865, Feb. 2022.

[43] B. S. Abunasser, M. R. J. Al-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, "Breast cancer detection and classification using deep learning xception algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 7, pp. 223–228, 2022.

[44] H. Yu and B. M. Wilamowski, "Levenberg–Marquardt training," in *Intelligent Systems*. Boca Raton, FL, USA: CRC Press, 2018, pp. 1–12.

[45] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014, doi: 10.1109/TBME.2014.2303852.

[46] Q. Qi, Y. Li, J. Wang, H. Zheng, Y. Huang, X. Ding, and G. K. Rohde, "Label-efficient breast cancer histopathological image classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2108–2116, Sep. 2019, doi: 10.1109/JBHI.2018.2885134.

[47] L. Liu, Y. Wang, P. Zhang, H. Qiao, T. Sun, H. Zhang, X. Xu, and H. Shang, "Collaborative transfer network for multi-classification of breast cancer histopathological images," *IEEE J. Biomed. Health Informat.*, vol. 28, no. 1, pp. 110–121, Jan. 2024, doi: 10.1109/JBHI.2023.3283042.

[48] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016, doi: 10.1109/TMI.2015.2458702.

[49] H. M. E. Misilmani, T. Naous, S. K. A. Khatib, and K. Y. Kabalan, "A survey on antenna designs for breast cancer detection using microwave imaging," *IEEE Access*, vol. 8, pp. 102570–102594, 2020.

[50] J. Li, B. Wang, D. Zhang, C. Li, Y. Zhu, Y. Zou, B. Chen, T. Wu, and X. Wang, "A preclinical system prototype for focused microwave breast hyperthermia guided by compressive thermoacoustic tomography," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2289–2300, Jul. 2021.

SOUMYA SARA KOSHY received the B.Tech. degree in computer science and engineering from Mahatma Gandhi University, Kottayam, Kerala, in 2008, and the M.Tech. degree in computer science and engineering from Calicut University, Kerala, in 2013. She is currently pursuing the Ph.D. degree with Vellore Institute of Technology, Chennai. Her research interests include medical image processing and deep learning.

L. JANI ANBARASI received the B.E. degree in computer science and engineering from Manonmaniam Sundaranar University, in 2000, and the M.E. and Ph.D. degrees from Anna University, in 2005 and 2015, respectively. She is currently an Assistant Professor (Sr.) with the School of Computing Science and Engineering, VIT Chennai. She has around 15 years of experience in various institutions. Her area of expertise includes cryptography, image processing, and soft computing techniques. She has published around 102 technical publications in various international journals and conferences. Her professional membership includes India Society for Technical Education (Life Member) and ACM. She received the Best Paper Award from the International Conference on Knowledge-Based Computing Technologies, Chennai, in February 2017, and the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), in December 2017.

• • •