Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

Breast Cancer Histopathological Image Classification Using Augmentation Based on Optimized Deep ResNet-152 Structure

R.Reenadevi[1*], T.Sathiya[2], B.Sathiyabhama[3]

[1,2,3]Department of Computer Science and Engineering, Sona College of Technology, Salem, India
*reenadevi@sonatech.ac.in

ABSTRACT

Breast cancer is the world's second leading malignant tumor, which is susceptible toboth men and women,butit'sfarmorecommoninwomen.Multipledatamining, machinelearning and deep learningtechniques were developed andapplied for classification. Despite significant improvements in medicine, pathological image analysis remains the most common method for diagnosing breast cancer.In pathological image analysis, Computer-Aided Diagnosis (CAD) is commonly used to assist pathologists improve diagnosis efficiency, accuracy, and consistency.Deep learning strategies have been investigated in recent studies to boost the effectiveness of pathological CAD. In this proposed work, an optimized deep ResNet structure is employed to extract more affluent and finercharacteristics from clinical images. The proposed technique is evaluated on the publicly availableBreakHis dataset. The results indicated that the proposed model in all the magnification levels outperforms the baseline techniques significantly.

Keywords: Breast Cancer, Histopathological, ResNet-152, BrakHis, Classification, CAD

Introduction

Breast cancer is one of the most debilitating and dangerous diseases that women face today, although early identification improves survival rates significantly.The precise diagnosis of breast cancer, on the other hand, is a challenging task.The American Cancer Society (ACS) estimates that there are more than 3.7 million breast cancer patients in the United States [1].A woman's possibility of dying from breast cancer is about 2.8%.Several tests, including mammograms, ultrasound images, MRIs, and biopsy samples, are used to detect breast cancer. In this,histopathology plays an importantrole in the investigation, since it is the gold standard for differentiating between benign and malignant tissue, as well as patients with in situ and invasive cancer [2].The significant goal of cancer treatment is to cure the disease, followed by life extension and pain alleviation when cure is not possible owing to advanced disease.

Currently, 35% of all types of cancers are regularlytreated successfully.When possible, treatment should result in a cure, after which the quality of life can be considered satisfactory. Symptom alleviation may occur as a result of curative treatment, but when cure is not possible, rapid symptom relief becomes critical.The current remedial for breast cancer is not a guaranteed cure and could cause extreme discomfort andhardshipforaffectedpatients.Mammography [3] and histopathologyarethemostcommonlyusedbreast cancer screening techniques to detect breast lesions. Even for physicians, it becomes difficult to correctly diagnose and what they do is simply to refer their patients for biopsy because of noisy images, they have their own confusions about lesions. Normally, it takes more time to consume the dataset, if this is benign or malignant. Neural network classification can do a better job classifying the data. So the residual network is a classic neural network used as a backbone for many detection tasks. It is necessary to solve the problem of identifying patches without noise.

The proposed CAD system improves the accuracy of interpreting mammograms by providing an important opinion to the radiologist. The proposed system could helpradiologists

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

to distinguish between normal and abnormal cases. The abnormalcasesarefurthersubdividedintomass or Micro-calcifications(MCs) [4];theabnormalitiesof images are further classified either as benign or malignant in the subsequent stages. The goal of this research work is to propose an effectivetechniquefor extracting thesalient texture features for classification and then give the classification results to evaluate the corresponding features, thereby developing a novel technique for histopathologicalclassification.

Themainobjectiveofthis workistodesignaneffectiveandautomatedCADschemeforthe classification ofimages.

Stage 1: To classify normal and abnormal lesions in breast tumors
Stage 2: To Classify the abnormal lesions into mass and MCs
Stage 3: To Classify benign or malignant based on the abnormal severity in mass and MCs

Related Work

For the recognition of breast cancer,Saxena et al. [5] implemented a CAD system by combining a kernelized extreme learning algorithm with ResNet50. They have used BisQue and BreakHis datasets for their analysis.They have resolved the issue of class disparity. Li et al. [6] have created a new, densely squeezed, and excited neural net (CNN) architecture.BreakHis dataset was used for assessing the architecture's performance. Wang et al. [7] combined CNN with a capsule network to create architecture for obtaining discriminate features.The authors employed histopathology data to classify breast cancer using deep feature fusion and routing.Boumaraf et al. [8] developed a transfer learning-based system for categorizing pathological images.They used a strategy of blocking sophisticated tuning and global networking standardization techniques.For efficient classification, the ResNet-18 deep neural network was utilized.

Toğaçar et al. [9] implemented BreastNet, a CNN-based architecture. The augmentation technique has been used to process and transfer the image. They used the BreakHis dataset for their analysis and obtained an accuracy of 98.8%.The study of machine learning (ML) and transfer learning methodologies was proposed by Sharma and Mehra [10].The handcrafted features are evaluated using ML algorithms. To evaluate the global features, transfer learning approaches are applied. They discovered that combining VGG16 with the Support Vector Machine (SVM) [25] produces the betteroutcome for all magnification factors of histopathological images. Burcak et al. [11] created a deep CNN technique for detecting breast cancer. They used various gradient algorithms to calculate the network's initial weight and update model parameters to improve learning and achieve better results.

Sharma and Mehra [12] designed a layer-by-layer fine-tuning technique to improve classification performance in a variety of medical care domains. They can also help with the selection of suitable layers from pre-trained networks.They fine-tuned the network using a transfer learning methodology and histopathology images for analysis.Asare et al. [13] used an augmented-based CNN approach to classify histopathological images.The model is built from scratch and minimizes overfitting. The authors used four deep learning optimization algorithms and demonstrated that the CNN model outperformed the other models with an accuracy of 89.92%.Carvalho et al. [14] used histopathological images to implement a CAD system to diagnose breast cancer. They extracted texture attributes from phylogenetic indexes. The extracted feature information is used to classify the data into four categories.

Materials and Methods

Theproposedsystem's work flowis shown in **Figure**

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

1whichincludescollectionofthebenigngroupandmalignant groupdatafromBreakHisdataset [15].Thecollecteddataispre-processedandstandardized to obtainimportant features. With the aid of ResNet-152 [16], pre-training and efficient feature set selection from the retrieved features are accomplished, and the selected features are categorized into one of two groups: benign and malignant.

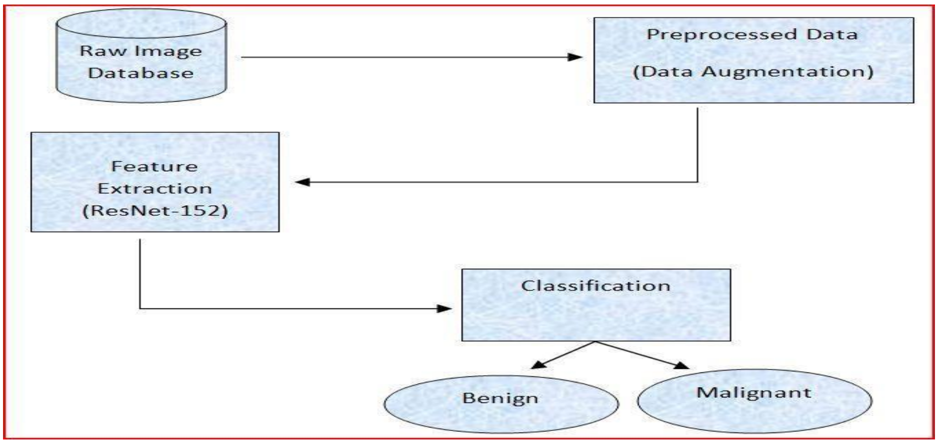

Figure 1. Proposed CAD System flow

BreakHis Dataset

The Breast Cancer Histopathological Image Classification (BreakHis) [15] is made up of 7,909 microscopic images of breast tumor tissue obtained and magnified by 400X, 200X, 100X, and 40X from 82 patients.This includes 5,429 malignant, and 2,480 benign with 700X460 pixels samples, RGB 3-channel samples, 8-bit depth on each channel of PNG format.The P & D Laboratory - Pathological Anatomy and Cytopathology, Brazil, collaborated on the creation of this database.

Preprocessing

Pre-processing [16] is a term used to describe manipulationsof images at the most basic abstractions for both input and output images.It is explicitly used to improve image data by suppressing undesired distortion or enhancing certain visual qualities that are relevant for later processing.Data Augmentation [17] is a technique for obtaining more images that employs random horizontal flipping, resize cropping, and rotation. The way the results of the final data processing are interpreted could be influenced by image pre-processing.

Image Augmentation

Image augmentation is a strong approach for creating variations in obtainable images so that an obtained image data set can be expanded [17].This generates creative and new images from an obtained image dataset that contains a wide range of possibilities. Deep learning Networks(CNN)[11] needsahugenumberofimages to be appropriately learned.This contributes to the performance of the model through better generalization and reduction of the over fit.The proposed ResNet-152may classify objects in different sizes, directions or illumination due to its variable property and this is shown in Table 1. As a result, augmentation [17] produces a wide variety of images for the classification, detection, or segmentation of images from a small set of images.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

Table 1 Augmentation Operations

Operations	Description
Randomrotation	The image is rotated and preserved.The rotations provide an intuitive, optional enhancement to generate a unique coordinate system.
Randomflip	The image is flipped so that an object as its mirror image can be identified. The flipping often used here is horizontal flipping.
Randomcrop	Cropping serves as a regulator, allowing us to construct a finer detector by usingneighboring regions as negative instances.

FeatureExtraction using ResNet-152

Residual networks, a type of neural network, have been proposed for extraction of features.ResNet[18] includes a residual learning unit to overcome the deterioration of deep-neural networks.This unit is designed as a feed forward network with a shortcut link that allows new inputs and outputs to be added.The key advantage is that it improves classification accuracy without adding to the model's complexity. This creates the ResNet -152 layer [19] by combining more three-layer blocks.The 152-layer ResNet has a less complex architecture than the 34-layer network VGG-16/19.The links between residual blocks were of significant benefit to the ResNet[18] architecture's residual connections.It maintains the information gained through training and enhances model building by raising the network's capacity.The following **Figure 2** shows the singleresidual block structure.

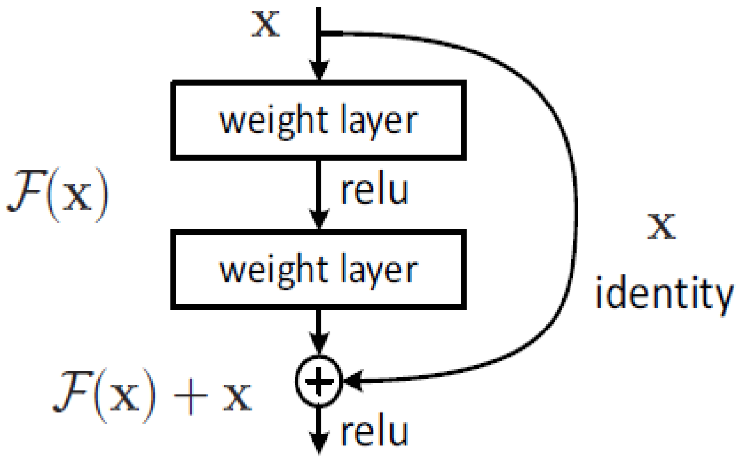

Figure 2. Single residual block

Classification

Classification is asupervised learning [21] [22] technique which dividesbreast cancer into categories by various criteria. Classification [20][23] in machine learning is the identification of which groups belong to the new instance based on training data that includes instances known to their class association.The histopathological type, the tumor grade, the tumor stage, expression of proteins and genes are the four major categories. Inthiswork,binary classification is done which will determine whether the image dataset is either benign ormalignant.

Deep Residual Network

ResNet employs skip connections to fit input from one layer to the next without altering it [18].Skip connection allows for a deeper network as illustrated in Figure 3. They commonly have convolution layers for typical deep learning networks that are connected, for classification tasks such as VGGNet,ZFNet, and AlexNet.The vanishing gradient problem occurs whenever the networksare deeper.A shortcut connection is introduced to the input x

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

following several convolution layers to resolve this problem.Shortcut carries out identity mapping, with additional zero padding and no extra parameters. All shortcuts are projections. This is used mostly to increase dimensions instead of identity. F(x) is any mapping that allows two layers of weight to fit R(x),

$$F(x)= R(x)+x \qquad (1)$$

R(x) denotes a residual mapping with regard to identity. If identity were optimal, it would be simple to set weights to zero.It is easier to find little fluctuations if the optimum mapping is closer to identity.Allplain or residualnetsaretrainedfromscratch and usestandard hyper-parameters to augmentation, and normalize batches. Deep ResNets can be trained without any difficulties and it is presented in **Figure 4**.

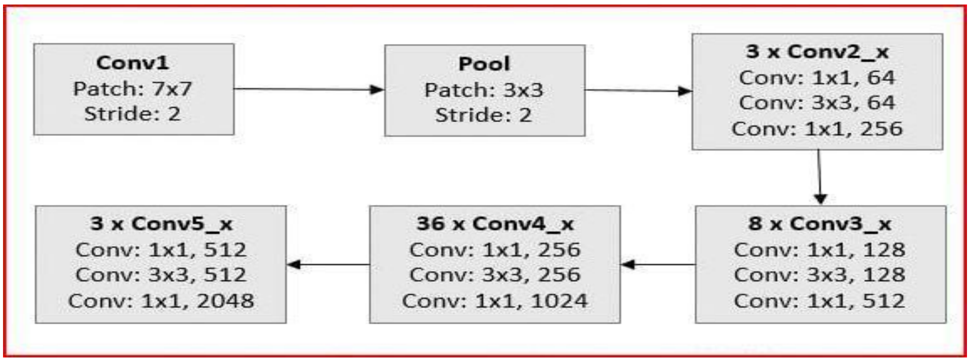

Figure 3. Deep Residual Network

Deeper ResNets have lower training errors, and alsorepresentation, optimization, and generalization ability with the following advantages:

(1) Allow models to go deeper
(2) It allows for very smooth forward and backward propagation, making it much easier to optimize deeper models
(3) Good deeper address generalization

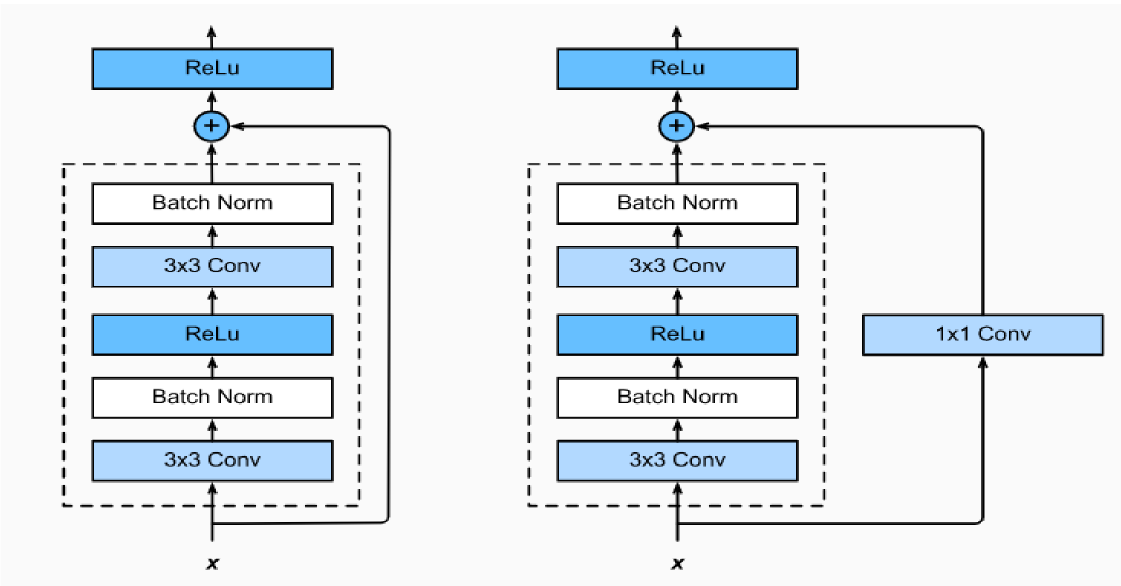

Figure 4. Left: regular ResNet block; Right: ResNet block with 1x1 convolution

Experimental Results

The proposed system is implemented by using NVIDIA Tesla K40 GPUs with Tensorflow.Histopathological database BreaKHis dataset [15] that includes 8 sub-

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

categories of breast cancers is used for analysis. The data set consists of 2 groups: benign tumors and malignant ones as shown in Table 2.Both breast tumors can be classified by pathologists based on the microscope aspect of tumor cells as illustrated in **Figure 5**.It will be categorize into 4 types of benign breast tumors: tubular adenona, adenosis, phyllodes tumor, and fibroadenoma; and 4 malignant tumors: carcinoma, papillary, lobular, and mucinous carcinoma.

Table 2.BreakHis Dataset

Magnification	Total	Benign	Malignant
400x	1820	588	1232
200x	2013	623	1390
100x	2081	644	1437
40x	1995	652	1370
Total	**7909**	**2480**	**5429**

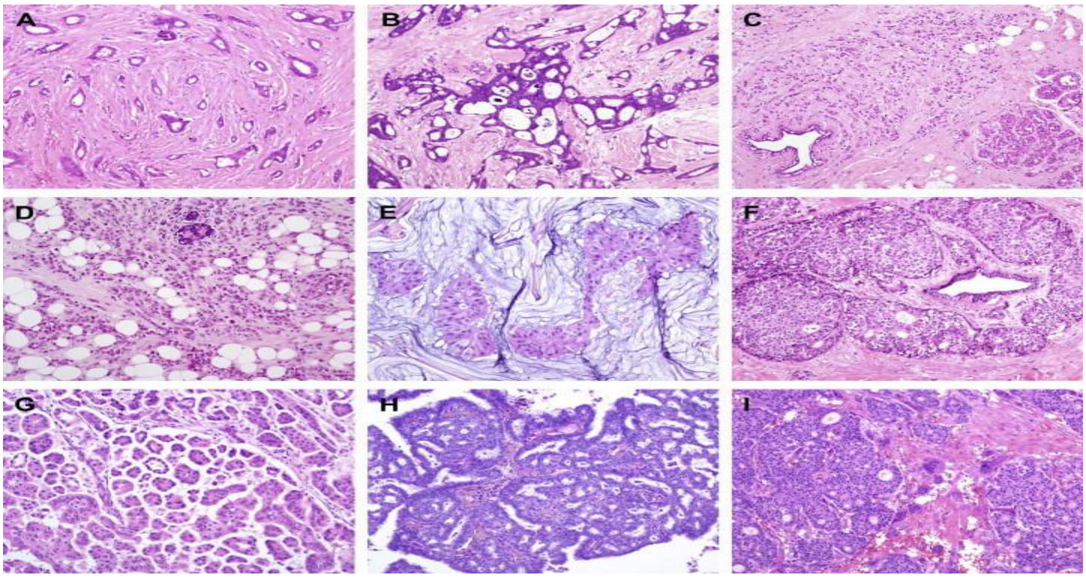

Figure 5. Sample BreakHis datasetImages

In the testing phase, data augmentation [17] with 90 degree rotations and horizontal flipping is chosen.The result obtained is evaluatedforspecificity, accuracy, sensitivity, precision, F1-score, and recall. These measures are associatedwith the rate of false-negative (FN), true negative (TN), true positive (TP), and false positive (FP) respectively. These metrics are expressed inthe following equations 2 to 6.

$$\text{Accuracy} = (TN + TP)/(FP + TN + FN + TP) \tag{2}$$

$$\text{Specificity} = TN/(TN + FP) \tag{3}$$

$$\text{Precision} = TP/(FP + TP) \tag{4}$$

$$\text{Recall} = TP/(FN + TP) \tag{5}$$

$$\text{F1 - score} = 2 * ((\text{recall} * \text{precision})/(\text{precision} + \text{recall})) \tag{6}$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

The following **Table 3** presents the performance comparison of the deep learning models VGGNet, ResNet and ResNet-152. It is clearly shown that the proposed augmented ResNet-152 outperforms the compared models and gives effective results.

Table 3 Performance comparison of the various Network model

Model	Magnification	Accuracy	Precision	Recall	F1-score
VGGNet	40x	92.2	92	94	93
	100x	93.3	96	94.2	95
	200x	95.1	96	96.2	96
	400x	94	93	95.7	94
ResNet	40x	96	95.8	95.5	95.6
	100x	94.5	96.5	94.9	95.7
	200x	93	93.6	97.8	96.4
	400x	92	95.5	94.4	96.6
Proposed work (ResNet152)	40x	97	96.5	97.1	96.8
	100x	98.6	97.4	96.6	97
	200x	97	97.5	97.1	97.2
	400x	96	97.2	96.7	97

The visualization of the performance comparison is demonstrated in following Fig.6. It is demonstrated that the proposed ResNet-152 having higher rates for all magnification factors compared to other methods.

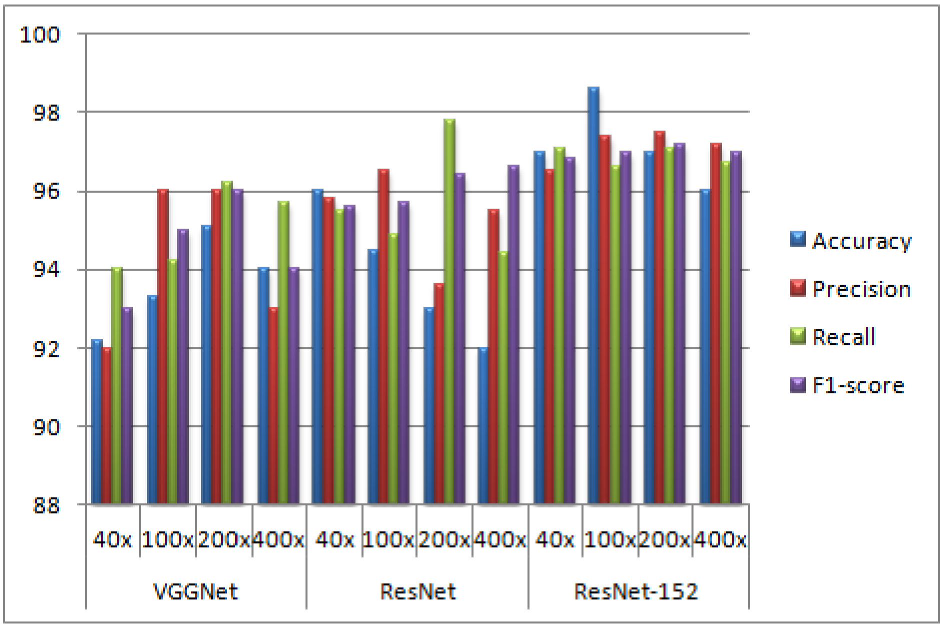

Figure 6. Visualization of performance comparison

Conclusion

The proposed deep learning technique ResNet-152 is a useful and responsible approach in comparison to traditional methods.This research also focused on detecting cancer subtypes.In this work, ResNet-152 deep learning structure is used for feature extraction and classification. Compared to other network models, deeper ResNets have lower training errors, optimization, and generalization ability. It allows for very smooth

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 5866 - 5874
Received 25 April 2021; Accepted 08 May 2021.

forward and backward propagation, making it much easier to optimize deeper models, and deeper address generalization. The empirical results demonstrated that ResNet-152 with augmentation gives better results compared to VGGNet, ResNetat all the magnification levels.In the future, other kinds of deep learning techniques will be considered and improved for breast cancer diagnosis.The proposed framework could be extended to different medical applicationsfor thediagnosis of pathological images.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249.

[2] Bahl, M., Mercaldo, S., McCarthy, A. M., & Lehman, C. D. (2021). Imaging surveillance of breast cancer survivors with digital mammography versus digital breast tomosynthesis. *Radiology*, *298*(2), 308-316.

[3] Reenadevi, R., Sathiya, T., &Sathiyabhama, B. (2021). Classification of Digital Mammogram Images using Wrapper based Chaotic Crow Search Optimization Algorithm. *Annals of the Romanian Society for Cell Biology*, 2970-2979.

[4] Sharma, S., Mehra, R., & Kumar, S. (2021). Optimised CNN in conjunction with efficient pooling strategy for the multi-classification of breast cancer. *IET Image Processing*, *15*(4), 936-946.

[5] Saxena, S., Shukla, S., &Gyanchandani, M. (2021). Breast cancer histopathology image classification using kernelized weighted extreme learning machine. *International Journal of Imaging Systems and Technology*, *31*(1), 168-179.

[6] Li, X., Shen, X., Zhou, Y., Wang, X., & Li, T. Q. (2020). Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PloS one*, *15*(5), e0232127.

[7] Wang, P., Wang, J., Li, Y., Li, P., Li, L., & Jiang, M. (2021). Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing. *Biomedical Signal Processing and Control*, *65*, 102341.

[8] Boumaraf, S., Liu, X., Zheng, Z., Ma, X., &Ferkous, C. (2021). A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, *63*, 102192.

[9] Toğaçar, M., Özkurt, K. B., Ergen, B., &Cömert, Z. (2020). BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, *545*, 123592.

[10] Sharma, S., &Mehra, R. (2020). Conventional machine learning and deep learning approach for
multi-classification of breast cancer histopathology images—a comparative insight. *Journal of digital imaging*, *33*(3), 632-654.

[11] Burcak, K. C., Baykan, Ö. K., &Uğuz, H. (2021). A new deep convolutional neural network model for classifying breast cancer histopathological images and thehyper parameter optimisation of the proposed model. *The Journal ofSupercomputing*, *77*(1), 973-989.

[12] Sharma, S., &Mehra, R. (2020). Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image. *The Visual Computer*, *36*(9), 1755-1769.

[13] Asare, S. K., You, F., &Nartey, O. T. (2020). Efficient, ultra-facile breast cancer histopathologicalimages classification approach utilizing deep learning optimizers.*International Journal of Computer Applications*, *11*, 9.

[14]Carvalho, E. D., Antonio Filho, O. C., Silva, R. R., Araujo, F. H., Diniz, J. O., Silva, A. C., ...

&Gattass, M. (2020). Breast cancer diagnosis from histopathologicalimages using textural features and CBIR. *Artificial Intelligence in Medicine*, *105*,101845.

[15] Spanhol, F. A., Oliveira, L. S., Petitjean, C., &Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, *63*(7), 1455-1462.

[16]Tripathy, S., &Swarnkar, T. (2020). A Comparative Analysis on FilteringTechniques Used inPreprocessing of Mammogram Image. In *Advanced Computinand Intelligen Engineering* (pp. 455-464). Springer,Singapore.

[17]Saini, M., & Susan, S. (2020). Deep transfer with minority data augmentation forimbalanced breast cancer dataset. *Applied Soft Computing*, *97*, 106759.

[18] e Silva, D. C. S., & Cortes, O. A. C. (2020). On Convolutional Neural Networks and Transfer Learning for Classifying Breast Cancer on Histopathological Images UsingGPU. In *CongressoBrasileirode EngenhariaBiomédica* (pp. 1-6).

[19] Chapala, H., &Sujatha, B. (2020, July). ResNet: Detection of Invasive Ductal Carcinoma in Breast Histopathology Images Using Deep Learning. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 60-67).IEEE.

[20] Sathiya, T., Reenadevi, R., &Sathiyabhama, B. (2021). Lung nodule classification inCT images using Grey Wolf Optimization algorithm. *Annals of the Romanian Societyfor Cell Biology*, *25*(6), 1495-1511.

[21] Sathiya, T., Reenadevi, R., &Sathiyabhama, B. (2021). Random Forest Classifierbased detection of Parkinson's disease. *Annals of the Romanian Society for Cell Biology*, 2980-2987.

[22] Sathiya, T., &Sathiyabhama, B. (2019). Fuzzy relevance vector machine based classification of lung nodules in computed tomography images. *International Journal of Imaging Systems and Technology*, *29*(3), 360-373.

[23] Sathiyabhama, B., Jayanthi, J., Sathiya, T., Ilavarasi, A. K., Kumar, S. U., Yuvarajan, V., &Gopikrishna, K. (2020). A novel Feature Selection Framework based on Grey Wolf Optimizer for Mammogram Image Analysis. Journal of Neural Computing and Applications.

[24] Sathiyabhama, B., Jayanthi, J., Sathiya, T., Ilavarasi, A. K., Kumar, S. U., &Yuvarajan, V. (2020). A grey wolf optimization for feature subset selection in the classification of breast cancer data.Journal of Soft Computing.

[25] Rajeswari, C., Sathiyabhama, B., Devendiran, S., &Manivannan, K. (2014). Bearingfault diagnosis using wavelet packet transform, hybrid PSO and support vector machine. Procedia Engineering, 97, 1772-1783.

[26] Shukla, A., Kalnoor, G., Kumar, A., Yuvaraj, N., Manikandan, R., &Ramkumar, M. (2021). Improved recognition rate of different material category using convolutional neural networks. Materials Today: Proceedings. Published.https://doi.org/10.1016/j.matpr.2021.04.307

[27] Mishra, A., Jain, H., Biswas, P., Thowseaf, S., &Manikandan, R. (2021). Integrated solution for optimal generation operation efficiency through dynamic economic dispatch on Software Technological Park of India. Materials Today: Proceedings. Published.https://doi.org/10.1016/j.matpr.2021.05.019