

Received February 8, 2021, accepted March 26, 2021, date of publication April 5, 2021, date of current version April 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071297

Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram

SHARAF J. MALEBARY¹, (Member, IEEE), AND ARSHAD HASHMI²

¹Department of Information Technology, Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdulaziz University, Jeddah 21911, Saudi Arabia

²Department of Information Systems, Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdulaziz University, Jeddah 21911, Saudi Arabia

Corresponding author: Arshad Hashmi (ahsyed@kau.edu.sa)

This work was supported by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant G: 321-830-1441.

ABSTRACT In recent years, deep learning techniques are employed in the mammography processing field to reduce radiologists' costs. Existing breast mass classification systems are implemented using deep learning technologies such as a Convolutional Neural Network (CNN). CNN based systems have attained higher performance than the machine learning-based systems in the classification task of mammography images, but a few issues still exist. Some of these issues are; ignorance of semantic features, analysis limitation to the current patch of images, lost patches in less contrast mammography images, and ambiguity in segmentation. These issues lead to increased false information about patches of mammography image, computational cost, decisions based on current patches, and not recovering the variance of patches intensity. In turn, breast mass classification systems based on convolutional neural networks produced unsatisfactory classification accuracy. To resolve these issues and improve the accuracy of classification on low contrast images, we propose a novel Breast Mass Classification system named BMC. It has improved architecture based on a combination of k- mean clustering, Long Short-Term Memory network of Recurrent Neural Network (RNN), CNN, random forest, boosting techniques to classify the breast mass into benign, malignant, and normal. Further, the proposed BMC system is compared with existing classification systems using two publicly available datasets of mammographic images. Proposed BMC system achieves the sensitivity, specificity, F-measure, and accuracy for the DDSM dataset is 0.97%, 0.98%, 0.97%, 0.96% and for the MIAS dataset is 0.97%, 0.97%, 0.98%, and 0.95% respectively. Further Area Under Curve (AUC) rate of the proposed BMC system lies between 0.94% - 0.97% for DDSM and 0.94%-0.98% for the MIAS dataset. The BMC method worked comparably better than other mammography classification schemes that have previously been invented. Moreover, the Confidence interval statistical test is also employed to determine the classification accuracy of the BMC system using different configurations and neural network parameters.

INDEX TERMS Deep learning, convolutional neural network, random forest, breast mass classification, recurrent neural network, boosting, machine learning, mammograms, medical image processing.

I. INTRODUCTION

Breast cancer is one of the most exceedingly invasive malignancies tumors that occurs in women and rarely in men [1]. It is considered the worst cancer after lung cancer due to the higher death rate in women [2]. According to the study of NCI (National Institute of Cancer) of china, from the year 2008 to 2012, 1.67 million cases of breast cancer are

reported and 522,000 fatalities [3]. Fortunately, it is also observed from the study of the last five years; if breast cancer is detected in the early stages, then the survival rate may reach more than 90% [4]. On the contrary, late diagnosis can cause it to spread to the lymph node and other body parts, which drops the survival rate by 27% [5]. Therefore, it is necessary to detect breast cancer at an early stage to save women's lives. The breast has a very complicated structure; sometimes, the specialist cannot identify the local lesions, and massive reading procedures also mislead the

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya¹.

diagnosis [6]. The diagnose procedure for breast cancer requires medical screenings. Computer-Aided Diagnosis (CAD) techniques based on medical screening is convenient and reduces the doctor's workload [7]. CAD has also fetched technical innovations in medical diagnosis. The popular CAD-based medical screenings are mammography, Magnetic Resonance Imaging (MRI), and ultrasound. Mammography is the most effective, reliable, has low cost and less harmful screening for early detection of breast cancer [8]. Automatic CAD with mammography has a higher accuracy rate of detection, operation convenient speeds up the diagnostic procedure, and saves the medical resources [9]. Further, the breast mass plays a distinctive sign in breast cancer diagnosis. The marginal information of the breast mass shows the biological characteristics and growth patterns [10]. The irregular margins of breast mass are often malignant tumors, and masses segmentation accuracy affects the Malignant Breast Masses Classification (MBMC). Therefore, masses classification is an essential procedure in the CAD of breast cancer. It helps the doctors in treating and diagnosing breast cancer in the early stages. The breast mass has various features such as illness boundaries, different shapes, and sizes. Therefore, accurate segmentation classification is challenging and popular a problem in CAD techniques [11].

Motivation: Many researchers have been working on the Breast Mass Classification (BMC) and developed various algorithms to extract various potential features from the mammographic images [12] using Machine Learning (ML) and Deep Learning (DL) techniques. Usually, the CAD system has four main parts: image preprocessing, lesion segmentation, feature extraction, and classification. Fig.1. Further, ML techniques have become a controversial issue for BMC in the medical image processing study [13]. This is because traditional ML techniques introduce the data sparsity issue [14]. Although traditional ML techniques used a handmade feature extraction method, it is very difficult to design an effective feature extraction technique due to sample diversity [15]–[20]. From the past few years, the DL technique has gained wide attention from researchers in various fields due to the capability of automatic feature removal, extract DL is widely used in bioinformatics [21]–[23].

CNN technology captures the relationship between input and output with non-linear mapping. You may also use several features combined in a single program. Using feature points will make changes much more significant than could ever be achieved by building from scratch. CNN technique has been applied to various Biomedical Image Processing (BIP) problems and segmentation of the medical images in one of them [24]. CNN technique has been continuously achieving good results in the mass segmentation process on public datasets. In the mass segmentation process, CNN utilized the image block's pixel as the input of the network for classification. It causes huge capacity overhead and low estimation efficiency due to repeated calculation of convolution during prediction and training [25].

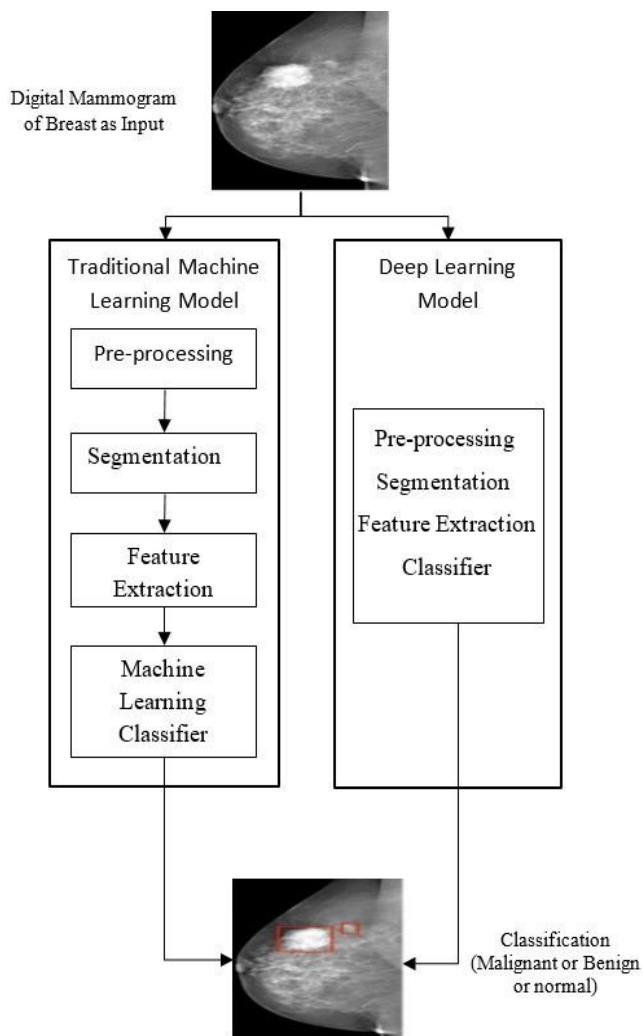


FIGURE 1. Graphical representation of difference between traditional machine learning based CAD system and deep learning based CAD system.

Additionally, the big sized image blocks require many pooling layers; due to this, segmentation accuracy is reduced. On the other hand, the utilization of a small-sized image block as input reduces receptive fields' size. These small-sized receptive fields only extract the local features and have limited segmentation accuracy. Previous researches also applied the CNN on the mammography-processing field [26]. Although the existing CAD system based on the CNN technique has attained higher performance than ML based on the CAD system in mammographic images' classification task, few issues are still existing. These issues are described below:

- Not able to find patches in less contrast mammography images. CNN technique reduces the weights prediction in generalizing for all segments. It only analyzes the current patches of images and sometimes makes decisions according to current patches.
- CNN does not find the relation between different patches of images, so the patch intensity variance does not recover by CNN.

- Variance generalization cannot be found by CNN that increases the false information about patches of mammography image.
- CNN finds the non-linearity between features on the current block; It is not linked to previous blocks. In turn, the model was refined iteratively and increased the cost of computation.

A. CONTRIBUTION

A novel BMC system is proposed with improved architecture based on the combination of Long Short-Term Memory network of Recurrent Neural Network (RNN- LSTM), CNN, random forest, boosting techniques, and mammographic images to resolve the issues as mentioned earlier for BMC. The proposed BMC system aims to classify the breast mass into three categories. These categories are benign, malignant, and normal. Further, the proposed BMC system is compared with existing classification systems using two publicly available datasets of mammographic images. The detailed contribution is described below:

- A new RNN-LSTM model is proposed to find the generalized feature weights.
- In the proposed BMC system, CNN model is integrated to extract the efficient low-level features.
- Transfer Learning method is also introduced to fine-tune the pre-trained CNN model for mass classification.
- To increase the classification performance of breast mass systems, reduce the variance and generalization error, an ensemble learning method is introduced in the proposed BMC using Random Forest and Boosting ML techniques.
- The detailed experimental results on two large published datasets, such as Mammographic Image Analysis Society MIAS) and Digital Database for Screening Mammography (DDSM), are presented.
- The proposed BMC system's performance is validated using five performance metrics AUC, sensitivity, specificity, F-measure, and accuracy.

The rest of this article is organized as follows. Section II has the related work for the illustration of the breast mass classification scheme. Section III has the details of the proposed breast mass classification (BMC) system. Experimental results and discussions are elaborated in section IV. Section V has the threats to validity, and finally, in section VI, conclusions are drawn.

II. RELATED WORK

This work aims to increase the performance of BMC using mammographic images. This section presents the recent works of CNN based on CAD systems for BMC.

To classify mass mammograms, Dhungel *et al.* [26] proposed the two structured prediction approaches named Structured Support Vector Machine (SVSM) and Conditional Random Field (CRF) in 2015. Both proposed approaches use potential functions based on deep convolution and belief networks. The results showed that the CRF approach

outperforms the SSVM in terms of training and inference time. In 2015, Carniero *et al.* [27] studied the CNN in medical images and proposed a new model for continuously extracting the features from the same full breast images. Initially, the CNN model is trained for each segmentation map, MLO, and CC views by utilizing the Imagenet pre-trained framework, and important features are then extracted. These extracted features are used in diagnosing breast cancer by calculating the data system and breast imaging reporting scores of DDSM and INBreast datasets. The proposed model generated the ROC of 0.9 over the surface. Samala *et al.* [28] proposed a CAD system using deep CNN with TL from mammograms. The proposed CAD system aims to detect the mass lesion in Digital Breast Tomosynthesis (DBT) by learning the image patterns from mammograms. The proposed system used 2282 digital mammograms and 324 DBT as a dataset. Proposed deep CNN based on the CAD system is compared with the existing feature-based system for the reduction of false-positive using ROC and AUC parameters. The proposed deep CNN based on the CAD system outperforms the feature-based CAD system with a difference of < 0.05 p-value.

To classify the benign and malignant tumors in Medio-Lateral Oblique (MLO) and Cranio-Caudal (CC) view, Bekker *et al.* [29] developed a new Siamese Neural Network (SNN) structure method based on CNN. The proposed method consisted of two neural networks and a view-level of decisions that are accomplished by a single neural layer. This neural layer combined the decisions of view level into a global decision for helping in biopsy results. The results showed that the proposed SNN based model improves the accuracy of the DDSM dataset when compared to existing standard NN based models. In 2016, Dhungel *et al.* [30] proposed a detection model based on RF and DL techniques. The proposed detection model is divided into three stages. In the first stage, the classifier is developed using Deep Belief Network (DBN) and the Gaussian Mixture Method (GMM) to generate the candidates. These candidates are fed into a cascade of two CNN for extracting the features. At the last stage, two RF classifiers are employed on these extracted features to reduce the False Positive (FP) and increase the detection system's True Positive (TP) rate. The proposed detection model achieved 0.96 ± 0.03 TP at 1.2 FP on the INBreast dataset and 0.75 TP at 4.8 FP on the DDSM-BCRP dataset as per image. Lévy and Jain [31] designed a new classification model using CNN for classifying the mammograms into benign and malignant. Further, to overcome the issue of limited training samples, the combination of data augmentation, transfer learning, and careful pre-processing is used in the proposed model. The proposed model achieved better results than the existing model on the DDSM dataset. To classify the mammogram images into two classes of breast density, Mohamed *et al.* [32] proposed a new model using the CNN technique. The proposed image classification model is tested on 22,000 images. Further, Area Under Curve (AUC) and Receiver Operating Characteristic (ROC) performance parameters are utilized to validate the performance.

The results showed that the accuracy of the proposed system increases with the size of training images.

To reduce medical specialists' fee cost, Jaffer [22] implemented a new model for detecting cancer automatically in mammograms using the CNN technique with SVM. Initially, images are resized and fed into CNN for feature extraction; finally, the SVM classifier is employed on extracted features for classifying the images. The proposed model outperforms the existing model by achieving 93.35% sensitivity and 93% accuracy on DDSM and MIAS datasets. To classify the breast tissue mammograms into normal and abnormal, Gardezi *et al.* [21] presented a new approach using the CNN technique. The proposed method used VGG-16 CNN with 3×3 convolutional filters for extracting the feature matrix of the IRMA dataset. After that, various ML classifiers are used to classify the mammograms using these feature matrixes' such as K-Nearest Neighbor (KNN), Binary Tree (BT), and SVM. The classification accuracy and AUC value of the proposed approach are 100% and 1.0, respectively.

Antropova *et al.* [23] proposed a new model to detect Breast Lesion Malignancy (BLM) in three medical imaging methods such as ultrasound, DCE-MRI, and digital mammography. The proposed model extracted two types of features. Initially, the low and mid-level features are extracted using CNN. After that, radiomic features are computed using fuzzy c-means. At last, both features are combined, fed into SVM and radial bases function classifier for classifying suspected lesion malignancy. The evaluation result showed that the proposed model performs well on less image preprocessing, has accurate error estimation, and computationally efficient. Chougrad *et al.* [33] proposed a mass lesion mammography classification system using CNN to help the medical specialist. Further, the author utilized the Transfer Learning (TL) and fine-tuned approach during CNN's training phase. The proposed system gained the 97.35% and 95.50% accuracy rate and 0.98 and 0.97% AUC rate on DDSM and BCDR datasets.

To classify the abnormalities of mammographic images of the mini MIAS dataset, Li *et al.* [34] proposed four classification models using the CNN technique named as CNN-2, CNN-4, CNN-2d, and CNN-4d. The various image preprocessing approaches are used, such as balancing, Local Histogram Equalization (LHE), Global Contrast Normalization (GCN), cropping, and augmentation. The performance of the four models is validated using accuracy, specificity, and sensitivity parameters. The performance of CNN-4d is better than other proposed models and achieved 90.63% sensitivity, 89.05% accuracy, and 87.67% specificity. The results of this work showed that CNN produces good results in diagnosing the medical images automatically. Rampun *et al.* [35] developed a new image classification model using the CNN technique. Initially, the proposed classification model is trained with epochs. After that best three CNN models are selected based on validation accuracy. Finally, the prediction results of the selected CNN model are combined to get the final prediction result. The experimental result revealed that the

accuracy of classification is increased by combining the selected CNN models' results. To identify the mass region automatically in mammography images, Diniz *et al.* [36] proposed a computational methodology using the CNN framework. The proposed computational methodology consists of two stages such as training and testing. In the training stage, the two models are initially developed to classify the breast tissues into non-dense and dense, and the region of the breast into non-mass and mass. The testing stage consists of further seven sub-stages such as segmentation, preprocessing, etc. to classify the breast region into non-mass and mass of the DDSM dataset. The proposed computational methodology gained 95.6% and 97.72% accuracy for non-dense and dense classes, respectively. To observe the stromal changes in breast cancer, Bejnordi *et al.* [37] implemented a new model using the deep CNN technique. Further, the proposed model successfully extracted stromal features and classified stroma around invasive cancer and benign biopsies in breast cancer. To improve the accuracy of the CNN classification model's using a small training dataset, Perre *et al.* [38] employed the transfer learning technique on three pre-trained CNN on the Imagenet dataset. This pre-trained CNN's are used to extract specific features. After that, extracted features and handcrafted features are fed into the SVM classifier to classify the lesions. To classify the malignant and benign mass in mammogram images, Ragab *et al.* [39] developed a new Breast Cancer Classification Model (BCCM) using a deep CNN model such as AlexNet and two segmentation methods. The first segmentation method manually identified the Region Of Interest (ROI), and the second method utilized the region and threshold approaches.

Further, AlexNet is used to extract the features, and these features are fed into SVM classifier for better classification accuracy. The proposed BCCM achieved higher accuracy and AUC than previous work on DDSM and CBIS-DDSM dataset and achieved higher. The accuracy of BCCM is 71.01% for DDSM and 87.2% for CBIS-DDSM. The AUC of BCCM is 88% for DDSM and 94% for CBIS-DDSM. To enhance the classification accuracy of mammographic images, Sun *et al.* [40] proposed a new multi-view CNN based on various mammography views. The detailed convolution of the proposed model automatically extracted the features from CC and MLO views.

Further, the penalty term is introduced into the loss function, such as cross-entropy, for reducing the misclassification rate. The proposed model is compared with various exiting models using MIAS and DDSM datasets and achieved better results. To enhance the performance of breast mass segmentation (BMS), Agnes *et al.* [41] proposed an automatic method by combining the U-net with AGs (attention gates). The proposed model contained a decoder that has U-net with AGs and an encoder that has a convolution network. The proposed model's performance is validated using sensitivity, F1-score, accuracy, mean intersection, and specificity parameters on the DDSM dataset. The proposed model integrated with U-net with AGs enhanced accuracy compared to the

previous model, U-net, Dense Net, and attention-based models. Ha *et al.* [43] developed a prediction model using CNN architecture to predict the atypical ductal hyperplasia (ADH) in mammographic images. The proposed CNN model has 15 hidden layers, five residual layers, and a 0.25 dropout rate. The proposed model's performance is validated using AUC, specificity, and sensitivity parameters on 298 images. The results revealed that CNN based model is capable of classifying the ductal carcinoma from ADH using mammographic images. Shu *et al.* [44] proposed an end-to-end the mammogram image classification model based on the CNN technique. This work aims to reduce the workload and cost of medical specialists. The author also proposed two structures of pooling. It is also observed that the proposed pooling structure improved the performance of CNN based classification model on the CBIS and INBreast dataset. To help the medical specialist efficiently detect breast cancer by accurately classifying the mammogram images, Agnes *et al.* [41] developed a multiscale all CNN (MA-CNN) approach utilizing the CNN framework. The proposed method extracted the specific features and automatically classified the mini-MIAS images into a benign, normal, and malignant category. The proposed MA-CNN approach enhanced the classification accuracy by extracting the considerable information utilizing multiresolution filters without affecting the CPU speed.

Further, the proposed MA-CNN approach achieved an overall 0.99 AUC and 96% sensitivity. To overcome the problem of existing CAD systems like overfitting and hand-crafted feature designing, Moon *et al.* [45] proposed a new CAD system. They combined the image fusion (IF) approach with the content representation of different images employing the different CNN frameworks like DenseNet, VGGNet, and ResNet on the ultrasound signals (US). The author designed a new dataset of 1687 tumor images, 734 images are malignant, and the left is benign. The proposed CAD system successfully detected the relevant features for predicting the class of tumor.

Jarosik *et al.* [46] developed a new model using the CNN technique for classifying the breast mass on US and radiofrequency data. The proposed classification model processed the radio frequency of 2D patches and their samples of amplitude automatically. The proposed classification model achieved higher accuracy rather than a traditional classifier based on the Nakagami parameter. Singh *et al.* [47] designed a cGAN (conditional generative adversarial network) model to segment breast cancer within mammogram images. The aim of cGAN is to generate a binary mask.

Further, the generated mask is classified into four classes of tumor shapes such as round, irregular, oval, and lobular using the proposed CNN based shape descriptor. The proposed cGAN model is tested on the INBreast dataset and gained 87% IOU and 94% dice coefficient. Although, the proposed shape descriptor achieved an 80% accuracy rate on DDSM. López-Cabrera *et al.* [48] developed a new CAD system based on TL and CNN technique. The proposed CAD system classified the mini-MIAS dataset into three categories that are malignant, benign, and normal. The overall accuracy of

the proposed CAD system is 86.05%. To check whether the DL-based models can transfer external data of mammograms with the various distribution of data, López-Cabrera *et al.* [48] study three existing CNN-based models and developed three new models based on the CNN technique. All six models' performance is tested on four datasets such as DDSM, MIAS, private data (UKy), and INBreast using the ROC performance parameter. The results revealed that DL based models could not successfully transfer the unseen external data and requires further validation and assessment. Shen *et al.* [49] proposed a new end-to-end system using the CNN approach to detect breast cancer. The result showed that the proposed system could be readily trained to achieve high accuracy on heterogeneous mammography and improved clinical tools to reduce false negative and positive screening mammography results. Vakanski *et al.* [50] proposed a new U-net architecture based on a deep learning model using attention blocks to focus on breast tumor segmentation's visual saliency. The author claimed that the proposed approach increased the robustness and accuracy. To automate the multimodal breast cancer classification system, Khan *et al.* [51] proposed a new system using the CNN model. The proposed automated system is tested on the BraTS datasets and achieved an accuracy of 97.8% for BraTs2015, 96.9% for BraTs2017, and 92.5% for BraTs2018, respectively. Shalol *et al.* [52] proposed a new system using the CNN model to classify tuberculosis in the chest. The proposed system is tested on Shenzhen and Dataset 2. The results showed that the proposed system outperformed the existing deep learning methods.

A. RESEARCH GAP

Based on overviewing recent related work, most of the researchers extracted low-level features and used end to end CNN based classification system.

As shown in Table 1, not a single researcher extracted the high level or semantic features from mammographic images. In this work, RNN-LSTM and CNN methods are proposed to extract the high and low-level features and to find the patches in less contrast mammography images. Very few researchers automatically extracted the ROI from the images. Therefore, in this work, ROI is extracted automatically through segmentation to reduce radiologist assistance costs. Although, a few pieces of research combined ML and DL techniques for BMC. Generally, DL based systems are nonlinear in nature and have a high variance that affected the final classification result. To enhance the classification performance of the BMC system, the DL based techniques (CNN and RNN-LSTM) are used to extract the specific features, and ML-based techniques (random forest and boosting) are utilized for the classification. The researchers do not target their work on feature sequence mapping. In this work, a sequence of features is extracted using the proposed RNN-LSTM to send current block mapping information on nonlinear activation functions. This process improves the fewer contrast images by mapping RNN-LSTM features with CNN features because both features combination can define previous block variance.

TABLE 1. Study of Breast Mass Classification (BMC) systems based on convolutional neural network.

Ref	Year	Aim	Methodology	Feature	Dataset	Limitation
[53]	2019	To detect breast cancer on screening mammograms	Develop a deep learning algorithm using the CNN model.	Low level Features	CBIS-DDSM and INBreast	Worked on ROI Annotation parameter tuning; it's not tuning other patches of image and not detect without ROI base images.
[54]	2020	Integrating visual saliency into a deep research model for breast tumor segmentation.	Develop new deep learning model using attention blocks and U-Net architecture.	High saliency features	Breast Ultrasound Images	Selected attention block after convolution and ignore others, which increased the probability of information loss.
[55]	2020	To automate multimodal tumor classification using the Deep learning method.	Develop new deep learning system using edge-based histogram equalization and discrete cosine transforms, VGG16 and VGG19 CNN model, correntropy-based joint learning, and partial least square approach.	Best and robust covariant features	BraTS	Used extreme learning, which increases the complexity and ignores the improvisation of segmentation
[56]	2020	To extract informative features from images for detecting tuberculosis from the chest.	Develop a novel hybrid method using Mobile Net, a CNN model, and Artificial Ecosystem-based Optimization.	Optimized features	Shenzhen and Dataset 2	Used the feature selection, which increases the information loss and feature selection not depend on convolution. So, reduce the non-linearity mapping efficiency of features.

If a variance is increased, then the block is rejected; in turn, noise is also reduced. Further, the work as mentioned earlier does not map the nonlinear feature on ensemble learning, but in this work, the nonlinear feature vectors produced by DL techniques are mapped on ensemble learning-based proposed model using Random Forest and Boosting to increase classification performance of breast mass, reduce the variance and generalization error.

III. PROPOSED BREAST MASS CLASSIFICATION (BMC) SYSTEM

In this section, a new BMC system is designed to classify the DDSM and MIAS mammogram images into normal, benign, and malignant using RNN-LSTM and CNN techniques with the ensemble learning method. Figure 2 demonstrates the basic structure of the BMC system. The planned BMC structure comprises five stages.

A. SEGMENTATION USING K-MEAN CLUSTER

Segmentation is a major processing stage in radiological diagnosis to study, understand, and direct treatments for mammographic photographs. The main objective of this process is to remove ROI to reduce the costs of radiologist assistance automatically. The pictures from the mammograms are taken as input. These images were then transformed into a series of ROIs or pixels. These ROIs are shown with the images labeled or masked. This process divides the image into different segments to process the important segments instead of the full-frame. The segmentation method has two modules in the proposed framework. The following modules are listed:

1) GROUP BY CLUSTERING

The ROI of the mammographic images is present in this subpart; the whole picture is segmented by the K-mean clustering algorithm [53]. This subpart is intended to produce the number of clusters. The centroid is initially measured. Each cluster point is then taken from the respective data points with the nearest centroid. The K-mean algorithm effectively adjusts the centroid using equation 1.

$$F = \sum_{x=1}^n \sum_{y=1}^k \|A_y^x - C_x\|^2 \quad (1)$$

where the F function is objective, n is cluster numbers, k is cases, A_y^x is case yth and C_x is cluster x centroid. The distance between the data is calculated with Euclidean points $\|A_y^x - C_x\|$. Image segmentation steps into the region by utilizing the k-mean clustering algorithm is described in Algorithm 1. This phase's output is clustered regions, and these regions are fed into the growing clustered region module.

2) GROWING THE CLUSTERED REGION

The clustered regions are taken as input in this phase. Firstly, the growing clustered region method generates the seed points by a centroid. After that, the neighboring pixels are examined; similarity is checked to these seed points. In this way, the clustered regions with similar intensity values are started growing. Further, mammogram images' pixels are classified into homogeneous sets, and regions of different breast tissue regions are segmented in mammograms

Algorithm 1 Group by Clustering

Input: Pixelized mammography images
 Output: Pixels are grouped together based on the similarity

- Step 1: Compute the Intensity distribution (pixel value)
- Step 2: Initialize the centroid by random
- Step 3: Initialize $\{u_i\} i^x = 1$
- Step 4: For each cluster C_x
- Step 5: Repeat: cluster the points based on their intensities using equation 2

$$C^{(y)} = \underset{x}{\operatorname{argmin}} \|A^{(y)} - u_y\|^2 \quad (2)$$

Step 6: Compute the new centroid for each of the cluster using equation 3

$$u_y = \frac{\sum_{y=1}^m 1 \cdot \{C_{(y)=x}\} A^{(y)}}{\sum_{y=1}^m 1 \cdot \{C_{(y)=x}\}} \quad (3)$$

where, y and j iterate over all the intensities and centroids respectively, u_y represents the centroid intensity.
 End for

Algorithm 2 Growing the cluster region

Input: Initialize a cluster or group using algo 1
 Output: Enhance clustering based on similarity

- Step 1: Input Image = Clustered Region (A, C).
 Generate seed point by centroid
- Step 2: Growing the region till intensity difference is greater than mean of all centroid
- Step 3: Add the neighbor pixels that are not already part of segmented area.
- Step 4: Calculate the mean of new region
- Step 5: Save (A, C)
- Step 6: Return to step 2.

as output. The steps of growing the clustered region are described in Algorithm 2.

B. PRE-TRAINED CNN MODEL

In this phase, CNN model-based feature extractor is designed as shown in Figure 2 using ResNet architecture and transfer learning method [54], [55]. The parameters of the CNN model (ResNet50) pertain to a natural image set known as ImageNet. It is utilized to extract the low-level features from the generated ROI of mammogram images. The structure of the CNN model in transfer learning consisted of four convolution layers (Conv) of ResNet architecture with a ReLU mapping activation function and a simple fully connected layer (FC), as shown in Table 2. The transfer learning method fine-tuned the pre-trained CNN model and generated the feature vectors for mass classification [56]. The output of the pre-trained CNN model is low-level features.

TABLE 2. CNN model for transfer learning.

Input : Image		
RESnet	Conv Block 1 (4*48*64) Pooling (2*24*64)	Conv + ReLU
		Patch 7*7
		Max Pool 3*3
	Conv Block 2 (2*24*64)	Conv + ReLU 1*1(64)
		Conv +ReLU 3*3 (256)
		Conv +ReLU 1*1 (512)
	Conv Block 3 (2*24*64)	Conv + ReLU 1*1 (128)
		Conv +ReLU 3*3 (128)
		Conv +ReLU 1*1 (512)
	Conv Block 4 (2*24*64)	Conv + ReLU 1*1 (256)
		Conv +ReLU 3*3 (256)
		Conv +ReLU1*1 (1024)
	Conv Block 5 (2*24*64)	Conv + ReLU 1*1 (512)
		Conv +ReLU 3*3 (512)
		Conv +ReLU1*1 (2048)
FC + ReLU (Dropout= 0.5)		
Output (Sigmoid):[0,1]		

C. RNN-LSTM MODEL

In this phase, the input mammogram image is fed into the RNN-LSTM model. The aim of this phase is to extract the high-level semantic features from the mammogram images.

LSTM is the backbone of our approach novelty because our main emphasis on semantic features, which collect by patch sequences, and LSTM makes an efficient sequence modeling. In our research sequence modeling, find the patches relation to particular class and group according to that we predict next patch pixel which depend on previous patch pixel.

$$F: P_i^{T+1} \rightarrow P_j^{T+2}$$

Here in P_i^{T+1} i^{th} pixel of ptch selected at $T+1$ and predicted j^{th} pixel at $T+2$

By this sequence modeling able to get semantic related patch based on temporal relation. By these relations get semantic features between patches. RNN-LSTM model can find the time sequence relationship between the image's pixels and the eigenvalue of the images. These eigenvalues are helpful in finding the importance of pixels at the temporal level. RNN-LSTM model also finds the temporal relationship

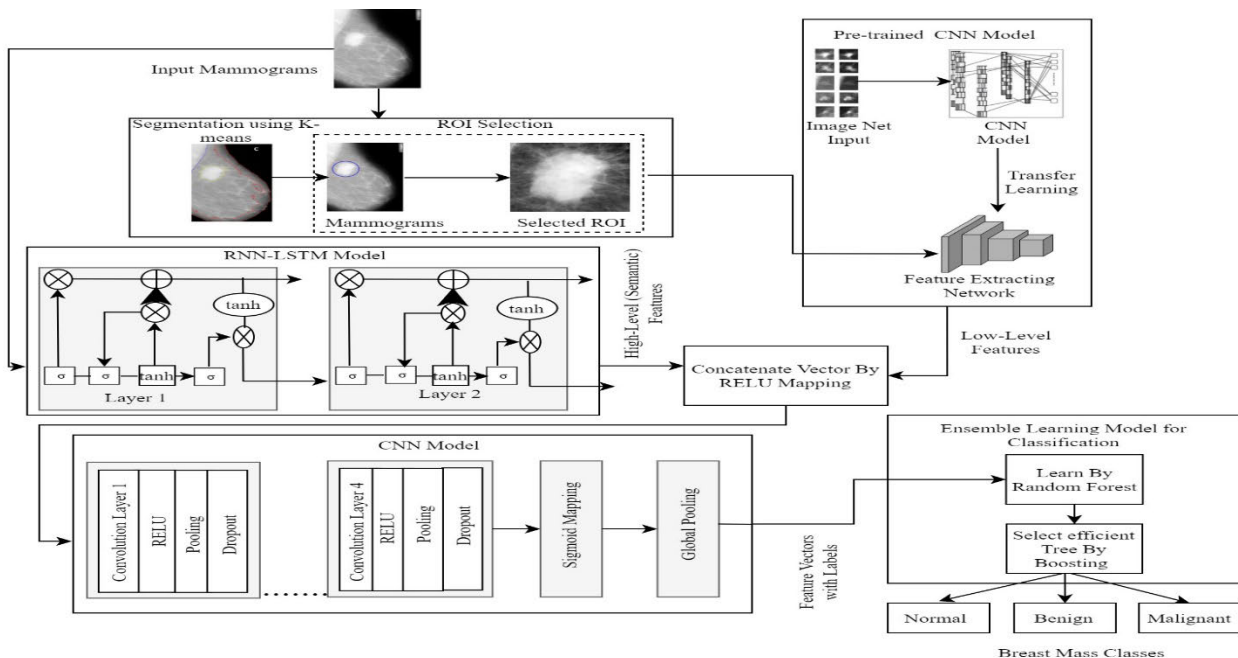


FIGURE 2. Architecture of proposed Breast Mass Classification (BMC) System.

between the features and generates feature vectors. We consider that these features are semantic. These features are related to each other by time or space. RNN-LSTM model gives the time-based relation so that we can say, extracted features are semantic. RNN is a type of neural network that used the previous output as the input while having the hidden states [52]. As represented in Figure.3., assume the RNN model takes the sequence of pixel $x = x_1, x_2, \dots, x_n$, produces the hidden states $H = H_1, H_2, \dots, H_n$ and output states $O = O_1, O_2, \dots, O_n$ in the following way:

$$O_t = \sigma(W_{H_t O_t} + b_t) \tag{4}$$

$$H_t = \sigma(W_{H_{t-1} H_t} H_{t-1} + W_{x_t H_t} x_t + b_{H_t}) \tag{5}$$

where, $W_{H_t O_t}$ represent the vector form the hidden unit H_t and output unit O_t , H_{t-1} represent the hidden unit for $t - 1$ pixel sequence, $W_{H_{t-1} H_t}$ is weight vector from the hidden unit H_{t-1} to H_t fir the sequence time t and b_{H_t}, b_t are the biases.

Further, LSTM stack is used to learn the time sequence features in the following way:

$$i_{g_t} = \tanh(W_{x_t i_{g_t}} x_t + W_{H_{t-1} i_{g_t}} H_{t-1} + b_{i_{g_t}}) \tag{6}$$

$$p_t = \sigma(W_{x_t p_t} x_t + W_{H_{t-1} p_t} H_{t-1} + b_{p_t}) \tag{7}$$

$$f_{g_t} = \sigma(W_{x_t f_{g_t}} x_t + W_{H_{t-1} f_{g_t}} H_{t-1} + b_{f_{g_t}}) \tag{8}$$

$$o_{p_t} = \sigma(W_{x_t o_{p_t}} x_t + W_{H_{t-1} o_{p_t}} H_{t-1} + b_{o_{p_t}}) \tag{9}$$

$$C_{e_t} = c_{e_{t-1}} \odot f_{g_t} + i_{g_t} \odot p_t \tag{10}$$

$$H_t = \tanh(C_{e_t}) \odot o_{p_t} \tag{11}$$

where, i_{g_t} represents the input gate, p_t indicates the prediction in starting layers, f_{g_t} represents the forget gate, H_t gives information of output, $b_{i_{g_t}}, b_p, b_{f_{g_t}}, b_{o_{p_t}}$ are the bias vectors, C_{e_t} indicates the state of cell and W_{xx} are the weight matrix.

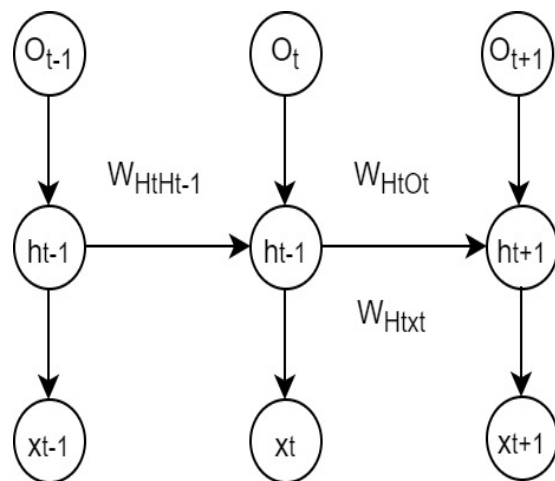


FIGURE 3. A graphical model for RN.

Both RNN and LSTM models are combined to extract the semantic features from the input mammogram images.

D. FEATURES CONCATENATION

In this phase, the features of CNN and RNN-LSTM models are combined. Both models are worked parallel for extracting the features from the input mammogram image.

The proposed system extracted two types of features, described as below:

- Low-level features that are extracted by the pre-trained CNN model.
- High level (semantic) features that are extracted by the RNN-LST model.

These features are concatenated by using the NLP approach based on the attention mechanism and the ReLU layer by

using equations 12 and 13.

$$a = \sigma(\tanh(W_{CNN} * X_{CNN} + W_{RNN} * f_{RNN})) \quad (12)$$

$$F = \max(a * X_{CNN}, a * X_{RNN}) \quad (13)$$

where X_{CNN} denotes the features of the medical image obtained by pre-trained CNN model, X_{RNN} is the gray medical image features obtained by the RNN-LSTM model, W_{CNN} and W_{RNN} are the weighted matrices, a represents the threshold of features weights after the feature concatenation and f is final weight matrix of the features. The dynamic features weights are obtained from equations 12 and 13. Further, the M feature matrix is generated using equation 14. This generated feature weight matrix acts as the input to the CNN model.

$$M = W * X \quad (14)$$

W feature weights are obtained from CNN, X feature weights are obtained from RNN, and M is the combined weight in equation (14)

E. CNN MODEL

In this phase, the CNN model is designed to extract the relevant features from the M feature matrix with weight. The detail about the CNN model is described in Figure 4. It has four convolution layers along with two max-pooling layers and one fully connected layer. The convolution layer contains many neurons. These neurons are connected with each other's and share biases and weights. This layer converts the matrix features into the feature map by utilizing convolution operations and using dropout to minimize the chances of overfitting. The key motivation of the Dropout layer is to reduce overfitting during training and testing, and thus the 0.27 value represents that only 27% of the data is retained. Although, the input feature matrix is mapped with kernel sets. These kernels produce the new feature map F_u and this procedure is known as convolution. The value of feature at location (s, d) in the u th feature map of k th layer, the value of C_u^k is computed by using equation 15.

$$C_u^k(s, d) = w_u^k I^k(s, d) + b_u^k \quad (15)$$

The max-pooling is used to reduce the computed feature maps dimension. ReLU activation function is used in each layer except the last output layer; for that, the sigmoid function is used. After the 4-convolution layer and 2 max-pooling operations, the output of the final convolution layer is propagated into the fully connected layer. Let the output of final convolution layer is E^g and it is calculated by using equation 16.

$$E^g = f(I^g) = f_{sigmoid}\left(\sum_u w_u * f(I^{g-1})\right) \quad (16)$$

where, u is the kernel numbers that are utilized in final convolution layer, w_u represents the kernel weight, $f(I^{g-1})$ is the value of activation function of $(g - 1)$ convolution layer. Finally, the fully connected layer contains the feature vectors with labels. These feature vectors are fed into ensemble learning model for breast mass classification.

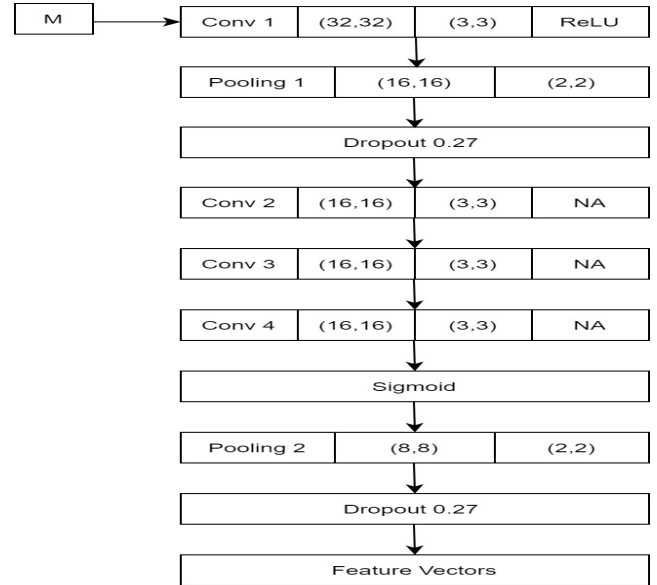


FIGURE 4. Parameter settings of CNN mode.

F. ENSEMBLE LEARNING MODEL FOR CLASSIFICATION

In this phase, the ensemble learning-based classification model is proposed by combining extreme gradient boosting with the random forest technique [30]. In the proposed system, the random forest is used to generate the different regression trees from the feature vectors with labels along with varying values of threshold. It is not capable of taking the final prediction decision regarding the different tree combinations. Hence, the ensemble learning model is proposed, as shown in Figure 5. In this, the random forest is utilized to generate all possible trees using feature space with respect to breast mass classes. Whereas the XG boosting technique is used to calculate the threshold values for model selection in the testing phase. It has an effective regularization function, in turn, less value of loss function is generated by using the following objective functions F :

$$F = \sum_{x=1}^n l(A_x, \hat{A}_x) + (\gamma T - (1 - \gamma) T) + \frac{1}{2} \lambda \sum_{y=1}^T w_y^2 \quad (17)$$

where, n indicates the given features with label instances and $\sum_{x=1}^n l(A_x, \hat{A}_x)$ represents the training loss function, which fits the data into L norm of leaf node by following equation:

$$L = (\gamma T - (1 - \gamma) T) + \frac{1}{2} \lambda \sum_{y=1}^t w_y^2 \quad (18)$$

All trees are built sequentially by using additive learning process. Each newly added tree learns from its former tree and update the prediction result by updating \hat{A}^{k-1} at k^{th} iteration. In this way, the input mammograms images are classified into normal, benign and malignant.

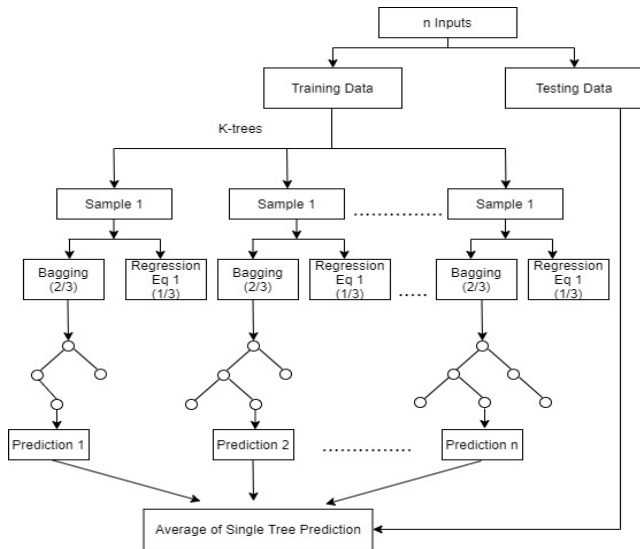


FIGURE 5. Graphical model for random forest with XG boosting.

G. TIME COMPLEXITY

In our research, three main parts of the CNN model, RNN-LSTM model, and Transfer parameters. In Transfer, parameters take O(1) complexity because of its predefined model. The matrix calculation in the CNN model takes O(n²), and the LSTM-RNN model also computes mostly on matrix, so it takes O(n²). So, the overall complexity of the proposed framework is O(1) + O(n²) + O(n²), and we can assume the complexity of the proposed framework is O(n²).

IV. EXPERIMENT RESULTS AND DISCUSSIONS

This section displays the experimental results of the proposed BMC system. Two publicly accessible datasets, such as DDSM and MIAS, use the proposed framework. F-metrics, precision, accuracy, and recall are used to determine the performance of the device proposed. To implement the process, the Windows 10 is enabled, an NVIDIA GeForce graphics card, a CPU @ 2.70GHz, and 16 GB RAM are used. In TensorFlow, Keras API is used to implement CNN and RNN-LSTM deep learning models [57].

A. DATASETS

Two datasets are utilized in this work, namely MIAS [58] and DDSM [59]. The mammogram images in both datasets are collected from 1994 to 2017 year. Both datasets are the largest mammography datasets available publicly and used widely to develop breast mass classification models.

- MIAS: This dataset is digitalized in nature and created in 1994 [58]. Further, this dataset contained 322 digital mammography films with the labeling of specialists at the abnormalities location that are collected from the United Kingdom National Breast Screening Program (UKNBSP). Images of this dataset are in low resolution.

- DDSM: Initially, this is constructed in 1999 and contained 10,480 digital film mammography images and 2,620 labeled cases [59]. ALL these cases are collected from two organizations- Wake Forest University School of Medicine (WFUSM) and Massachusetts General Hospital (MGH). These images (films) are non-uniformly scanned at various organizations with scanners and included optical density mapping and various gray levels of images. Further in this work subset (curated breast imaging) of DDSM is utilized for evaluating the performance of the proposed BMC system. This subset has 2620 samples; among them, 1445 samples are malignant, and 1175 are benign.

B. TRAINING

The proposed system’s objective is to extract the low and semantic features from the input mammogram image for classifying it into normal, benign, and malignant. This objective is formulated as a cross-entropy training loss function [60] that is expressed as below:

$$E = -\frac{1}{n} \sum_{i=1}^n [L^i \log L_o^i + (1 - L^i) \log (1 - L_o^i)] \quad (19)$$

where L_oⁱ represents the i-th neuron output result in the output layer, Lⁱ reflects the corresponding goal effects. The decent back spread and stochastic gradient [61] are used to minimize the loss of the samples. After minimizing the loss, the extracted features are learned by the random forest classifier. Further, a classifier is optimized by using equations 17 and 18. The objective function that is presented using equation 17, iteratively fits in equation 18. If it converges, then training is an exit, and the model is generated. The whole training process improved the features, features non-linear mapping, ensemble classifier, and also optimized an ensemble classifier. Additionally, the tuning parameters are refined in every step of different phases to make the model efficient.

Further, in this experiment ten-fold cross-validation method is utilized for dividing the dataset into testing and training dataset to evaluate the performance of proposed BMC system.

C. PERFORMANCE METRICS

Five efficiency measurements are used to quantify the performance of the proposed BMC method as shown below [20], [24].

Precision: Defines the appropriately labeled images for benign, malignant, and normal mammograms. The precision of the device proposed is determined by equation 20.

$$\text{Accuracy} = \frac{P + N}{P + N + F + G} \quad (20)$$

In equation 20, P denotes correctly classified positive samples, N represents correctly classified negative samples, F denotes incorrectly classified samples as correctly classified and G denotes correctly classified samples as incorrectly samples.

Specificity: It specifies the true negatives that can be correctly classified by diagnostic test. Specificity is computed using equation 21.

$$\text{Specificity} = \frac{P}{P + F} \tag{21}$$

Sensitivity: This metrics indicates actual patients who have disease among all patients. Sensitivity metric is computed using equation 22.

$$\text{Sensitivity} = \frac{P}{P + G} \tag{22}$$

F-measure: It is demonstrated as expression of specificity and sensitivity. It only takes the positive samples either classified as benign, malignant and normal. Equation 23 is used to compute the F-measure.

$$F - \text{measure} = \frac{2P}{2P + P + G} \tag{23}$$

AUC: It is presented as the ability of the classifier to differentiate between the benign, malignant and normal mammograms. It is used as a summary of the Receiver Operator Characteristics (ROC) curve.

D. RESULTS AND ANALYSIS

This subsection is the experimental findings of the BMC system. The proposed framework is evaluated with two separate datasets as specified above. Other tests, such as F-measure, Sensitivity, Specificity, AUC, and Accuracy, are used to determine these experimental results. The viability of the proposed method can be demonstrated on the minimum number of data and to guide further researchers in reducing the resources intensive process of obtaining labels. The training dataset is created by using the different numbers of patients, such as 20,30,40,50,60,70,80,100 from MIAS and DDSM datasets, for fine-tuning. The proposed BMC system consists of ResNet RNN-LSTM-CNN based network for feature extraction and random forest–boosting method for classification. The proposed system results are contrasted with another architecture system based on the ResNet-ResNet network and ResNet-VGG network [57] by using different training subsets. The proposed system’s experimental results are also compared with other mammography classification methods of existing breast cancer detection systems [26,29,33,43,44,61]. The confusion matrices of the proposed BMC system’s classification result are presented in Tables 3 and 4 for both datasets.

In both datasets, a total of 1000 images are shown in the confusion matrix, out of which 300 are normal,350 are Benign, and the rest 350 are malignant. It can be seen from confusion matrix tables that the malignant mammo-gram detection was not performed effectively because of the patches overlap. Benign mammograms show the maximum overlap in the DDSM dataset; thus, semantic feature selection improved the patches overlapping. The proposed BMC system showed better results on the DDSM dataset rather

TABLE 3. Confusion matrix of proposed BMC system using DDSM dataset.

Classes	Normal	Benign	Malignant
Normal	295	5	5
Benign	3	345	15
Malignant	2	5	330

TABLE 4. Confusion matrix of proposed BMC system using MIAS dataset.

Classes	Normal	Benign	Malignant
Normal	298	3	10
Benign	1	346	8
Malignant	1	1	332

than the MIAS dataset. Its main reason is the size of mam-mograms and blurriness. As compared to existing systems, the proposed BMC system performed significantly better in all classes.

The results of the proposed system are divided according to the MIAS and DDSM datasets as represented below:

1) RESULTS BASED ON DDSM DATASET

This subsection represents the experimental results of the proposed BMC system using the DDSM dataset. The exper-imental results of the proposed transfer learning model with ResNet RNN-LSTM-CNN based network are compared with ResNet-ResNet network and ResNet-VGG network-based models [57]. The AUC parameter is considered to evalu-ate the efficiency and effectiveness of transfer learning of the proposed BMC system. Table 5 and Figure 6 show the

TABLE 5. AUC Comparison of transfer learning efficiency of proposed and existing end to end systems with different training subsets of DDSM dataset.

Patients	Images	RESnet-RESnet	RESnet-VGG	Proposed
20	100	0.92	0.93	0.94
30	180	0.93	0.93	0.95
40	400	0.93	0.94	0.95
50	500	0.93	0.94	0.95
60	670	0.94	0.95	0.96
70	800	0.95	0.95	0.96
80	880	0.94	0.93	0.97
90	900	0.93	0.93	0.97
100	1000	0.92	0.94	0.97

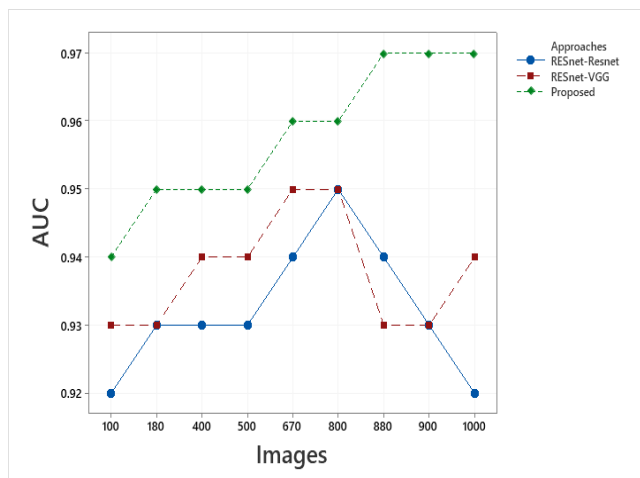


FIGURE 6. AUC comparison by number of images of proposed and existing systems using DDSM datasets.

AUC comparison results based on transfer learning efficiency of the proposed and the existing end to end systems (ResNet-ResNet and ResNet-VGG) with different training subsets of DDSM datasets.

It can be seen that there was a near-exact computation of the AUC for each training dataset. AUC rate of the proposed transfer learning model with ResNet RNN-LSTM-CNN based network lies between 0.94% and 0.97% with different patients and images. The ACU rate of ResNet-ResNet and ResNet-VGG network-based systems is between 0.92% to 0.95% and 0.93% to 0.95%, respectively. The graphical representation of three AUC measurements of the proposed BMC system and current systems using the DDSM dataset is provided in Figure 6. We use all these parameters in order to determine the efficiency of the BMC framework that we have proposed. The experimental results and comparisons of BMC systems and DDSM models are analyzed through the National Environment Agency (NEA) datasets. These systems are proposed by Bekker *et al.*, Dhungel *et al.*, Jaffer *et al.*, Sun *et al.*, Li *et al.* ResNet-ResNet and ResNet-VGG [26], [29], [22], [40], [41], [57]. Table 6 illustrate the comparison of different mammography classification methods on the DDSM dataset using different parameters.

It is observed that the BMC is able to calculate the f-measure, sensitivity, and specificity rates for each mammogram’s images of the DDSM dataset. It can be seen from Table 6, sensitivity, specificity, F-measure, and accuracy of the proposed system for the DDSM dataset is 0.97%, 0.98%, 0.97%, and 0.96%, respectively. It is also stated that the proposed BMC system achieves a better accuracy rate than other traditional mammography classification methods on the DDSM dataset. The accuracy rate of the proposed BMC system is 0.96 %. Whereas, the accuracy rate of Bekker *et al.*, Dhungel *et al.*, Jaffer *et al.*, Sun *et al.*, Agnes *et al.*, ResNet-ResNet, and ResNet-VGG systems are 79%, 90%, 93%, 0.82%, 0.83%, 0.89%, and

TABLE 6. Comparison of different mammography classification methods on DDSM dataset using different parameters.

Mammography Classification Methods	Sen.	Spec.	F	A	Y
Bekker <i>et al.</i> [29]	0.79	0.79	-	0.79	2016
Dhungel <i>et al.</i> [26]	-	-	-	0.90	2016
Jaffer <i>et al.</i> [22]	0.93	-	-	0.93	2017
Sun <i>et al.</i> [40]	-	-	-	0.82	2019
Agnes <i>et al.</i> [41]	-	-	-	0.83	2020
ResNet-ResNet [49]	0.91	0.93	0.93	0.89	2020
ResNet-VGG [57]	0.95	0.96	0.96	0.94	2020
Proposed	0.97	0.98	0.97	0.96	2021

Sen: Sensitivity, Spec: Specificity, F: F-measure, A: Accuracy, Y=Year

0.94% respectively. The sensitivity rate of Jaffer *et al.*, ResNet-ResNet and ResNet-VGG systems is 0.93%, 0.91%, and 0.95%, respectively. The specificity and F-measure rate of ResNet-ResNet and ResNet-VGG systems is 0.93%, 0.96%, 0.93% and 0.96% respectively. In conclusion, the classification and detection rate of the proposed BMC system is improved significantly than Bekker *et al.*, Dhungel *et al.*, Jaffer *et al.*, Sun *et al.*, Agnes *et al.*, ResNet-ResNet and ResNet-VGG systems on DDSM datasets. Further, the confidence interval (CI) of the classification result is also calculated using equation 24.

$$CI = z * \text{sqrt}((AUC * (1 - AUC))/n) \tag{24}$$

whereas, $z = 1.96$ and n is the ‘number of images’. The confidence interval of proposed BMC system is illustrated in Table 7 using different number of images and patients.

2) RESULTS BASED ON MIAS DATASET

This subsection provides an overview of the experimental results for the proposed BMC system when applied to MIAS dataset. Table 8 show the AUC comparative findings based on the efficiency of a system using a transfer learning approach and the accuracy of ResNet-VGG with different training subsets of MIAS datasets.

The proposed system explicitly calculates the AUC rates for each training dataset. AUC rate of the proposed transfer learning model with ResNet RNN-LSTM-CNN based network lies between 0.94 and 0.98 with different patients and images. The AUC rate of ResNet-ResNet and ResNet-VGG network-based systems are between 0.92 to 0.96 and 0.90 to 0.95, respectively.

TABLE 7. Confident interval of proposed BMC systems with different training subsets of DDSM dataset.

Patients	Images	Proposed	Confidence interval
20	100	0.94	0.004
30	180	0.95	0.003
40	400	0.95	0.002
50	500	0.95	0.002
60	670	0.96	0.002
70	800	0.96	0.001
80	880	0.97	0.001
90	900	0.97	0.001
100	1000	0.97	0.001

TABLE 8. AUC Comparison of transfer learning efficiency of Proposed and Existing methods with different training subsets of MIAS dataset.

Patients	Images	ResNet-ResNet	ResNet-VGG	Proposed
20	100	0.92	0.91	0.95
30	180	0.92	0.92	0.94
40	400	0.94	0.93	0.96
50	500	0.93	0.92	0.95
60	670	0.95	0.94	0.96
70	800	0.94	0.95	0.97
80	880	0.96	0.92	0.96
90	900	0.94	0.91	0.98
100	1000	0.93	0.90	0.97

Figure 7 shows a graphical depiction of AUC comparison levels of proposed BMC and current systems using photos representing discrepancies based on the MIAS dataset. We use all these parameters in order to determine the efficiency of the BMC framework that we have proposed. The experimental result will be compared with the MIAS dataset and other data to validate the BMCC framework. These systems are from Jaffer *et al.*, Sun *et al.*, Agnes *et al.*, ResNet-ResNet, and ResNet-VGG. This is shown in Table 8. It is obvious that the proposed BMC method accurately measures the precision, recall, f-measure, sensitivity, and specificity rates for each mammogram’s image in the MIAS dataset. It can be seen from Table 9, sensitivity, specificity, F-measure, and accuracy of the proposed system for the MIAS dataset are 0.97, 0.97, 0.98, and 0.95, respectively. It is also stated that the proposed BMC system achieves better accuracy, specificity, sensitivity, and F-measure rate

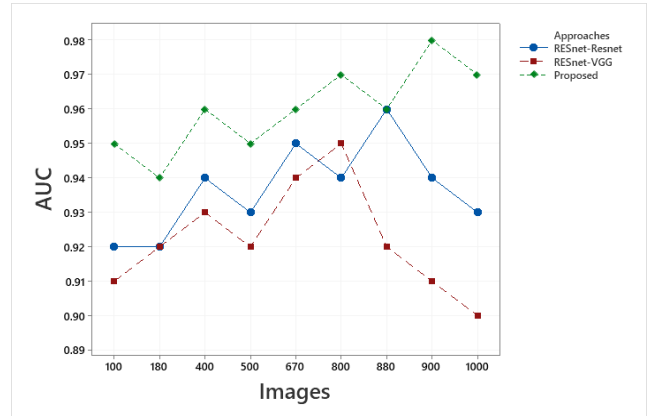


FIGURE 7. AUC comparison by number of images of proposed and existing systems using MIAS database.

TABLE 9. Comparison of different mammography classification methods on MIAS dataset using different parameters.

Mammography Classification Methods	Sen.	Spec.	F	A	Y
Jaffer et al. [22]	0.93	-	-	0.93	2017
Sun et al. [40]	-	-	-	0.82	2019
Agnes et al. [41]	-	-	-	0.83	2020
ResNet-ResNet [49]	0.87	0.89	0.91	0.92	2020
ResNet-VGG [57]	0.88	0.92	0.93	0.94	2020
Proposed	0.97	0.97	0.98	0.95	2021

Sen: Sensitivity, Spec: Specificity, F: F-measure, A: Accuracy, Y=Year

than other traditional mammography classification methods on the MIAS dataset. The accuracy rate of the proposed BMC system is 0.95, whereas Jaffer *et al.*, Sun *et al.*, Agnes *et al.*, ResNet-ResNet and ResNet-VGG have accuracy rates of 0.93, 0.82, 0.83, 0.92, and 0.94, respectively. Sensitivity ratings of Jaffer *et al.*, ResNet-ResNet, and ResNet-VGG systems are 0.93, 0.87, and 0.88, respectively. ResNet-ResNet and ResNet-VGG have specificity and F-measure rates of 0.89, 0.91, 0.92, and 0.93, respectively. This paper finds that both classification and detection rates increase significantly than Jaffer *et al.*, Sun *et al.*, Agnes *et al.*, ResNet-ResNet, and ResNet-VGG systems on the MIAS dataset. The confidence interval for the predictive performance of the proposed benchmark cohort method is shown in Table 10.

3) CONFIDENCE INTERVAL TEST

This subsection illustrates the statistical results of the predicted classification result of the proposed BMC system. The statistical tests are used to confirm the existence and validate the newly proposed BMC system. The aim of this tests is to determine the significance between results obtained using different configurations and parameters of neural networks. The significant confidence limit is also calculated from all accuracy results. The statistics of confidence and error are

TABLE 10. Confident interval of proposed BMC systems with different training subsets of MIAS dataset.

Patients	Images	Proposed	Confidence interval
20	100	0.95	0.01
30	180	0.94	0.010
40	400	0.96	0.007
50	500	0.95	0.006
60	670	0.96	0.005
70	800	0.97	0.005
80	880	0.96	0.004

demonstrated in Table 11 using different epoch, learning rates, filters, predicted accuracy, and the same activation function. The statistical results showed that the confidence interval of the highest predicted classification accuracy is 0.006398195. The classification error is 0.077 ± 0.0063 at 95 confidence and 97.5% accuracy. The error percentage, i.e., $7.7 \pm .63$.

4) DISCUSSION AND FUTURE DIRECTION

On both datasets, the proposed BMC system results demonstrate that the proposed BMC system attained a higher accuracy rate compared to existing systems with 20 patients and 100 images. The AUC rate of the proposed BMC system has approached the maximum as the size of the training subset increased on both the datasets. It is suggested from the evaluation results that the proposed BMC system require less dataset for learning and recognizing the textures and shapes of malignant, normal, and benign ROI's and for adjusting the different intensity profiles found in different datasets of mammography images. The approach of the proposed BMC system can be successfully utilized for fine-tuned the proposed ensemble learning-based image classifier using small training subsets with labels. This advantage can reduce the burden of creating the training set for various mammography platforms. It can also be seen from tables 6 and 9 that, proposed BMC system achieves higher accuracy, sensitivity, specificity, and F-measure on both datasets. These results clearly showed that the proposed BMC system successfully extracts low and semantic features from the images. An ensemble learning-based classifier learns these features and detects breast cancer efficiently. The proposed system also deals with blurriness of images by clustering the pixels, and these pixels are refined using transfer learning and convolution network. After that, these pixels make a sequence model between the patches by LSTM-RNN. These matches improve the pixel-wise feature mapping and its robustness in noisy images, but we cannot claim its 100% improvement but mapping improve to some extent. Future work could enhance the proposed model by using different angles of mammography images and patch wise classification along with reduced resources.

V. THREATS TO VALIDITY

In this work use, two benchmark datasets first is MIAS, and the second is DDSM. These datasets considering to evaluate

the performance of proposed and existing approaches. Each approach has two phases training and testing. These phases, validated by ten cross-validations and the proposed approach, get significant results on both datasets after simulation. Same simulation results apply to different approaches of deep learning, but some potential threats of validity will consider.

A. THREATS OF EXTERNAL VALIDITY

Threats to External Validity corresponds to the generalization of findings. In this work, two open-source datasets are used for mammography image classification. It is possible that industrial development and open source projects have certain differences. The industrial development project might require standard code quality. But we try to minimize the threat by considering MIAS and DDSM datasets. Both datasets have strong medical expertise background. Another threat is regarding the platform used to implement the projects. These data are generalized since they came from diverse populations. Hence, it is stated that threatens also associated with the external validity of our approach, as results obtained on a biased dataset are less generalizable.

B. THREATS OF INTERNAL VALIDITY

The threats to internal validity correspond to the selection of performance measure for evaluating the performance of the machine-learning classifiers. In this work, precision, recall, f-measure, AUC, and accuracy measures are adopted to assess the performance. But accuracy can be considered an important parameter to classify the mammography images in three classes. Several existing studies adopt these parameters to evaluate the performance of deep learning techniques for mammography image classification. But, sometimes, accuracy cannot lead towards a discrete conclusion. Because it can be described using the sum of correctly classified benign, malignant, and normal. In this paper, approaches depend on semantic and general features that further learn by different Convolution layers and generalize them. After that, optimize the learning process using random forest, which optimizes using boosting. Optimizations depend on decision trees, and decision trees depend on semantic and convolution generalize features.

C. THREATS OF STATISTICAL CONCLUSION VALIDITY

Threats to statistical conclusion validity correspond to the significant difference between the performances of deep learning techniques. To address the same, the Friedman test is adopted to determine the relationship between methods. It is a straightforward test and computed the ranking of each technique. Some statistics are calculated to check the significant difference between the methods. If the critical value is greater than the Friedman test value, hypothesis (Ho) is not rejected, and there is no significant difference between the performances of techniques. If rejected then, the method is significantly different than others. The proposed approach is substantially different from other techniques as a hypothesis (Ho) is rejected in our work. But this test does not consider

TABLE 11. Statistical analysis of predicted classification results of proposed BMC systems with different configurations and parameters of neural network.

Epoch	Learning rate	Activation	Filters	Accuracy	Error	Confidence
20	0.23	ReLU	32	0.92	0.08	90
20	0.34	ReLU	16	0.93	0.07	92
20	0.4	ReLU	8	0.94	0.06	93
40	0.23	ReLU	32	0.954	0.046	90
40	0.34	ReLU	16	0.934	0.066	89
40	0.4	ReLU	8	0.96	0.04	95
50	0.23	ReLU	32	0.967	0.033	96
50	0.34	ReLU	16	0.9675	0.0325	95
50	0.4	ReLU	8	0.975	0.025	95
80	0.23	ReLU	32	0.956	0.044	95
80	0.34	ReLU	16	0.967	0.033	95
80	0.4	ReLU	8	0.977	0.023	95
100	0.23	ReLU	32	0.9654	0.0346	97
100	0.34	ReLU	16	0.954	0.046	98
100	0.4	ReLU	8	0.934	0.066	95
120	0.23	ReLU	32	0.96	0.04	95
120	0.34	ReLU	16	0.95	0.05	95
120	0.4	ReLU	8	0.934	0.066	95
150	0.23	ReLU	32	0.923	0.077	95
150	0.34	ReLU	16	0.967	0.033	95
150	0.4	ReLU	8	0.9756	0.0244	95
200	0.23	ReLU	32	0.923	0.077	80
200	0.34	ReLU	16	0.967	0.033	80
200	0.4	ReLU	8	0.9567	0.0433	90

the size of datasets. The relative size of datasets is also an important parameter to determine whether the technique is significantly different than others or not. But this test is capable of choosing the best performer and worst performer. It also groups the methods that do not have a significant difference between performances.

VI. CONCLUSION

This paper proposes a new RNN-LSTM model to find the generalized feature weights and pass them to the convolution network. The proposed method is able to send information on current block mapping on nonlinear activation functions. Further, the proposed BMC system improved the less contrast by mapping RNN-LSTM features with CNN features because the combination of both features can define previous block variance. If the variance is increased, then the block is rejected; in turn, noise is also reduced. In the proposed BMC system, CNN extracted the efficient low-level features, and RNN-LSTM extracted the high-level features. The low-level features like segmentation, edge, color, texture, shape detection, and high-level features are known as semantic

features, such as the type of patch present in which segmented block or its pathology situations related to the patch.

The Transfer Learning method is also introduced to fine-tune the pre-trained CNN model for mass classification. To increase the classification performance of breast mass systems and reduce the variance and generalization error, an ensemble learning method is introduced in the proposed BMC using Random Forest and Boosting ML techniques.

The detailed experimental results on two large published datasets, such as Mammographic Image Analysis Society (MIAS) and Digital Database for Screening Mammography (DDSM), are presented. The performance of the proposed BMC system is validated using five performance metrics (AUC, sensitivity, specificity, F-measure, and accuracy). Proposed BMC system achieved the higher sensitivity, specificity, F-measure, and accuracy rate for the DDSM dataset that is 0.97%, 0.98%, 0.97%, 0.96% and for the MIAS dataset is 0.97%, 0.97%, 0.98%, and 0.95% respectively. Further Area Under Curve (AUC) rate of the proposed BMC system lies between 0.94% - 0.97% for DDSM and 0.94%-0.98% for the MIAS dataset.

The performance of the proposed BMC system outperforms the previously reported classification systems for mammographic images. The proposed BMC system reduced the ambiguity in segmentation and edges of tumors or patches, increased the efficiency of nonlinear mapping of features and classification accuracy of low contrast images. The confidence interval statistical test is also employed to determine the classification accuracy of the BMC system using different configurations and parameters of the neural network.

ACKNOWLEDGMENT

The authors acknowledge with thanks DSR for technical support.

REFERENCES

- [1] J. S. Whang, S. R. Baker, R. Patel, L. Luk, and A. Castro, "The causes of medical malpractice suits against radiologists in the United States," *Radiology*, vol. 266, no. 2, pp. 548–554, Feb. 2013, doi: [10.1148/radiol.12111119](https://doi.org/10.1148/radiol.12111119).
- [2] A. Felman, "What to know about breast cancer," *Med. News Today*, 2019, [Online]. Available: <https://www.medicalnewstoday.com/articles/37136>
- [3] W. Chen, R. Zheng, S. Zhang, H. Zeng, T. Zuo, C. Xia, Z. Yang, and J. He, "Cancer incidence and mortality in China in 2013: An analysis based on urbanization level," *Chin. J. Cancer Res.*, vol. 29, no. 1, pp. 1–10, 2017.
- [4] S. Xu, H. Wu, and R. Bie, "CXNet-m1: Anomaly detection on chest X-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2019, doi: [10.1109/access.2018.2885997](https://doi.org/10.1109/access.2018.2885997).
- [5] C. E. DeSantis, J. Ma, A. G. Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA, Cancer J. Clinicians*, vol. 67, no. 6, pp. 439–448, Nov. 2017, doi: [10.3322/caac.21412](https://doi.org/10.3322/caac.21412).
- [6] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, and K. Wang, "An efficient approach for automated mass segmentation and classification in mammograms," *J. Digit. Imag.*, vol. 28, no. 5, pp. 613–625, Oct. 2015.
- [7] O. Akin, S. B. Brennan, D. D. Dershaw, M. S. Ginsberg, M. J. Gollub, H. Schoder, D. M. Panicek, and H. Hricak, "Advances in oncologic imaging: Update on 5 common cancers," *CA, Cancer J. Clinicians*, vol. 62, no. 6, pp. 364–393, Oct. 2012.
- [8] H. Min, S. S. Chandra, N. Dhungel, S. Crozier, and A. P. Bradley, "Multi-scale mass segmentation for mammograms via cascaded random forests," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 113–117.
- [9] M. Dong, Z. Wang, C. Dong, X. Mu, and Y. Ma, "Classification of region of interest in mammograms using dual contourlet transform and improved KNN," *J. Sensors*, vol. 2017, pp. 1–15, Nov. 2017. [10.1155/2017/3213680](https://doi.org/10.1155/2017/3213680).
- [10] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, Mar. 2009, doi: [10.1109/itb.2008.2009441](https://doi.org/10.1109/itb.2008.2009441).
- [11] J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer-aided detection of breast masses on full field digital mammograms," *Med. Phys.*, vol. 32, no. 9, pp. 2827–2838, Aug. 2005, doi: [10.1118/1.1997327](https://doi.org/10.1118/1.1997327).
- [12] Y. Akhtar and D. P. Mukherjee, "Detection of architectural distortion from the ridges in a digitized mammogram," *Signal, Image Video Process.*, vol. 12, pp. 1285–1292, Mar. 2018. [10.1007/s11760-018-1281-1](https://doi.org/10.1007/s11760-018-1281-1).
- [13] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast cancer detection and diagnosis using mammographic data: Systematic review," *J. Med. Internet Res.*, vol. 21, no. 7, Jul. 2019, Art. no. e14464, doi: [10.2196/14464](https://doi.org/10.2196/14464).
- [14] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
- [15] Y. A. Reyad, M. A. Berbar, and M. Hussain, "Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification," *J. Med. Syst.*, vol. 38, no. 9, p. 100, Sep. 2014, doi: [10.1007/s10916-014-0100-7](https://doi.org/10.1007/s10916-014-0100-7).
- [16] P. S. Mukherjee, "A multi-resolution and adaptive 3-D image denoising framework with applications in medical imaging," *Signal, Image Video Process.*, vol. 11, no. 7, pp. 1379–1387, Oct. 2017.
- [17] Y.-Q. Song, Z. Liu, J.-M. Chen, F. Zhu, and C.-H. Xie, "Medical image segmentation based on non-parametric mixture models with spatial information," *Signal, Image Video Process.*, vol. 6, no. 4, pp. 569–578, Nov. 2012.
- [18] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: [10.1109/tpami.2002.1017623](https://doi.org/10.1109/tpami.2002.1017623).
- [19] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using K-Nearest neighbor machine learning algorithm," in *Proc. 9th Int. Conf. Develop. eSystems Eng. (DeSE)*, Aug. 2016, pp. 35–39.
- [20] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, Apr. 2020.
- [21] S. J. S. Gardezi, M. Awais, I. Faye, and F. Meriaudeau, "Mammogram classification using deep learning features," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 485–488.
- [22] M. A. Jaffar, "Deep learning based computer aided diagnosis system for breast mammograms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 7, pp. 286–290, 2017.
- [23] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.*, vol. 44, no. 10, pp. 5162–5171, Oct. 2017, doi: [10.1002/mp.12453](https://doi.org/10.1002/mp.12453).
- [24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [25] C. Tian, Y. Xu, L. Fei, and K. Yan, "Deep learning for image denoising: A survey," in *Proc. Int. Conf. Genet. Evol. Comput.*, 2018, pp. 563–572.
- [26] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 605–612.
- [27] G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Munich, Germany: Springer*, 2015, pp. 652–660. [Online]. Available: <https://www.events.tum.de/frontend/index.php?sub=22>, doi: [10.1007/978-3-319-24574-4_78](https://doi.org/10.1007/978-3-319-24574-4_78).
- [28] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Med. Phys.*, vol. 43, no. 12, pp. 6654–6666, Nov. 2016, doi: [10.1118/1.4967345](https://doi.org/10.1118/1.4967345).
- [29] A. J. Bekker, H. Greenspan, and J. Goldberger, "A multi-view deep learning architecture for classification of breast microcalcifications," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 726–730.
- [30] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2015, pp. 1–8.
- [31] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," 2016, *arXiv:1612.00542*. [Online]. Available: <http://arxiv.org/abs/1612.00542>
- [32] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz, and S. Wu, "A deep learning method for classifying mammographic breast density categories," *Med. Phys.*, vol. 45, no. 1, pp. 314–321, Jan. 2018, doi: [10.1002/mp.12683](https://doi.org/10.1002/mp.12683).
- [33] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Comput. Methods Programs Biomed.*, vol. 157, pp. 19–30, Apr. 2018, doi: [10.1016/j.cmpb.2018.01.011](https://doi.org/10.1016/j.cmpb.2018.01.011).
- [34] B. Li, Y. Ge, Y. Zhao, E. Guan, and W. Yan, "Benign and malignant mammographic image classification based on convolutional neural networks," in *Proc. 10th Int. Conf. Mach. Learn. Comput.*, Feb. 2018, pp. 247–251.
- [35] A. Rampun, B. W. Scotney, P. J. Morrow, and H. Wang, "Breast mass classification in mammograms using ensemble convolutional neural networks," in *Proc. IEEE 20th Int. Conf. E-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–6.

- [36] J. O. B. Diniz, P. H. B. Diniz, T. L. A. Valente, A. C. Silva, A. C. D. Paiva, and M. Gattass, "Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 1256, pp. 191–207, Mar. 2018, doi: [10.1016/j.cmpb.2018.01.007](https://doi.org/10.1016/j.cmpb.2018.01.007).
- [37] B. E. Bejnordi, M. Mullooly, R. M. Pfeiffer, S. Fan, P. M. Vacek, D. L. Weaver, S. Herschorn, L. A. Brinton, B. van Ginneken, N. Karssemeijer, A. H. Beck, G. L. Gierach, J. A. W. M. van der Laak, and M. E. Sherman, "Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies," *Mod. Pathol.*, vol. 31, no. 10, pp. 1502–1512, Oct. 2018.
- [38] A. C. Perre, L. A. Alexandre, and L. C. Freire, "Lesion classification in mammograms using convolutional neural networks and transfer learning," *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 7, nos. 5–6, pp. 550–556, Nov. 2019, doi: [10.1080/21681163.2018.1498392](https://doi.org/10.1080/21681163.2018.1498392).
- [39] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, p. e6201, Jan. 2019, doi: [10.7717/peerj.6201](https://doi.org/10.7717/peerj.6201).
- [40] L. Sun, J. Wang, Z. Hu, Y. Xu, and Z. Cui, "Multi-view convolutional neural networks for mammographic image classification," *IEEE Access*, vol. 7, pp. 126273–126282, 2019.
- [41] S. A. Agnes, J. Anitha, S. I. A. Pandian, and J. D. Peter, "Classification of mammogram images using multiscale all convolutional neural network (MA-CNN)," *J. Med. Syst.*, vol. 44, no. 1, p. 30, Jan. 2020, doi: [10.1007/s10916-019-1494-z](https://doi.org/10.1007/s10916-019-1494-z).
- [42] S. Li, M. Dong, G. Du, and X. Mu, "Attention dense-U-Net for automatic breast mass segmentation in digital mammogram," *IEEE Access*, vol. 7, pp. 59037–59047, 2019, doi: [10.1109/access.2019.2914873](https://doi.org/10.1109/access.2019.2914873).
- [43] R. Ha, S. Mutasa, E. P. V. Sant, J. Karcich, C. Chin, M. Z. Liu, and S. Jambawalikar, "Accuracy of distinguishing atypical ductal hyperplasia from ductal carcinoma *in situ* with convolutional neural network-based machine learning approach using mammographic image data," *Amer. J. Roentgenol.*, vol. 212, no. 5, pp. 1166–1171, May 2019, doi: [10.2214/ajr.18.20250](https://doi.org/10.2214/ajr.18.20250).
- [44] X. Shu, L. Zhang, Z. Wang, Q. Lv, and Z. Yi, "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2246–2255, Jun. 2020.
- [45] S. Li, M. Dong, G. Du, and X. Mu, "Attention dense-U-net for automatic breast mass segmentation in digital mammogram," *IEEE Access*, vol. 7, pp. 59037–59047, May 2019, doi: [10.1109/ACCESS.2019.2914873](https://doi.org/10.1109/ACCESS.2019.2914873).
- [46] P. Jarosik, Z. Klimonda, M. Lewandowski, and M. Byra, "Breast lesion classification based on ultrasonic radio-frequency signals using convolutional neural networks," *Biocybernetics Biomed. Eng.*, vol. 40, no. 3, pp. 977–986, Jul. 2020, doi: [10.1016/j.bbe.2020.04.002](https://doi.org/10.1016/j.bbe.2020.04.002).
- [47] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, and J. Torrents-Barrena, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112855, doi: [10.1016/j.eswa.2019.112855](https://doi.org/10.1016/j.eswa.2019.112855).
- [48] J. D. López-Cabrera, L. A. L. Rodríguez, and M. Pérez-Díaz, "Classification of breast cancer from digital mammography using deep learning," *Inteligencia Artif.*, vol. 23, no. 65, pp. 56–66, 2020.
- [49] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019.
- [50] A. Vakanski, M. Xian, and P. E. Freer, "Attention-enriched deep learning model for breast tumor segmentation in ultrasound images," *Ultrasound Med. Biol.*, vol. 46, no. 10, pp. 2819–2833, Oct. 2020.
- [51] M. A. Khan, I. Ashraf, M. Alhaisoni, R. Damasevicius, R. Scherer, A. Rehman, and S. A. C. Bukhari, "Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists," *Diagnostics*, vol. 10, no. 8, p. 565, Aug. 2020.
- [52] A. T. Sahlol, M. A. Elaziz, A. T. Jamal, R. Damasevicius, and O. F. Hassan, "A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features," *Symmetry*, vol. 12, no. 7, p. 1146, Jul. 2020.
- [53] X. Wang, G. Liang, Y. Zhang, H. Blanton, Z. Bessinger, and N. Jacobs, "Inconsistent performance of deep learning models on mammogram classification," *J. Amer. College Radiol.*, vol. 17, no. 6, pp. 796–803, Jun. 2020, doi: [10.1016/j.jacr.2020.01.006](https://doi.org/10.1016/j.jacr.2020.01.006).
- [54] W. Wiharto and E. Suryani, "The segmentation analysis of retinal image based on K-means algorithm for computer-aided diagnosis of hypertensive retinopathy," *Indonesian J. Electr. Eng. Informat.*, vol. 8, no. 2, pp. 419–426, 2020.
- [55] M. Farooq and A. Hafeez, "COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs," 2020, *arXiv:2003.14395*. [Online]. Available: <https://arxiv.org/abs/2003.14395>
- [56] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: [10.1109/jproc.2020.3004555](https://doi.org/10.1109/jproc.2020.3004555).
- [57] X. Yu, W. Pang, Q. Xu, and M. Liang, "Mammographic image classification with deep fusion learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.
- [58] J. Hung, A. Goodman, D. Ravel, S. C. P. Lopes, G. W. Rangel, O. A. Nery, B. Malleret, F. Nosten, M. V. G. Lacerda, M. U. Ferreira, L. Rénia, M. T. Duraisingh, F. T. M. Costa, M. Marti, and A. E. Carpenter, "Keras R-CNN: Library for cell detection in biological images using deep neural networks," *BMC Bioinf.*, vol. 21, no. 1, pp. 1–7, Dec. 2020, doi: [10.1186/s12859-020-03635-x](https://doi.org/10.1186/s12859-020-03635-x).
- [59] J. E. Oliveira, M. O. Gueld, A. D. A. Araujo, B. Ott, and T. M. Deserno, "Toward a standard reference database for computer-aided mammography," *Proc. SPIE*, vol. 6915, Mar. 2008, Art. no. 69151Y.
- [60] K. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, M. Sallam, K. Chang, and K. Woods, "The digital database for screening mammography," in *Proc. 3rd Int. Workshop Digit. Mammography*, vol. 58, 1996, p. 27.
- [61] A. Kukkar, R. Mohana, A. Nayyar, J. Kim, B.-G. Kang, and N. Chilamkurti, "A novel Deep-learning-based bug severity classification technique using convolutional neural networks and random forest with boosting," *Sensors*, vol. 19, no. 13, p. 2964, Jul. 2019, doi: [10.3390/s19132964](https://doi.org/10.3390/s19132964).



SHARAF J. MALEBARY (Member, IEEE) received the Ph.D. degree in computer science and engineering from the University of South Carolina, USA. He is currently an Assistant Professor and the Head of the IT Department, Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, Saudi Arabia. His research interests include machine learning, mobile, wearable, and ubiquitous technologies.



ARSHAD HASHMI received the Ph.D. degree in computer science from Shri J. J. T. University, Rajasthan, India, under the esteemed guidance of Prof. Dr. S. K. Yadav and Associate Prof. Dr. Amir Ahmad, College of Information Technology, United Arab Emirates University. He is currently an Assistant Professor with the Faculty of Computing and Information Technology in Rabigh and the Department of Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include machine learning, deep learning, network technologies, and the IoT.