IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 30, 2019, accepted November 9, 2019, date of publication November 13, 2019, date of current version November 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953318

Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network

HASAN NASIR KHAN[1,2], AHMAD RAZA SHAHID[1,2], BASIT RAZA[1,2], AMIR HANIF DAR[1,2], AND HANI ALQUHAYZ[3]

[1]Medical Imaging and Diagnostics Laboratory, National Centre of Artificial Intelligence, Islamabad 45550, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
[3]Department of Computer Science and Information, College of Science in Zulfi, Majmaah University, Al Majma'ah 11952, Saudi Arabia

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

The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. R-1441-31.

ABSTRACT Breast cancer is the second most common cause of cancer-related deaths among women. Early detection leads to better prognosis and saves lives. The 5-year survival rate of breast cancer is 99% if it is located only in breast. Conventional computer-aided diagnosis (CADx) systems for breast cancer use the single view information of mammograms to assist the radiologists. More recent work has focused on more than one views. Existing multi-view based CADx systems normally employ only two views namely Cranio-Caudal (CC) and Medio-Lateral-Oblique (MLO). The information fusion of the two views proved the effectiveness of the system for mammogram classification which cannot be achieved by single view information. However, combining the information of four views of mammograms increases the performance of classification. In this study, we propose Multi-View Feature Fusion (MVFF) based CADx system using feature fusion technique of four views for classification of mammogram. The complete CADx tool contains three stages, the first stage have the ability to classify mammogram into abnormal or normal, second stage is about classification of mass or calcification and in the final stage classification of malignant or benign classification is performed. Convolutional Neural Network (CNN) based feature extraction models operate on each view separately. These extracted features were fused into one final layer for ultimate prediction. Our proposed system is trained on four views of mammograms, after data augmentation. We performed our experiments on publicly available datasets such as CBIS-DDSM (Curated Breast Imaging Subset of DDSM) and mini-MIAS database of mammograms. In comparison with literature the MVFF based system is performed better than a single view-based system for mammogram classification. We have achieved area under ROC curve (AUC) of 0.932 for mass and calcification and 0.84 for malignant and benign, which is higher than all single-view based systems. The value of AUC for normal and abnormal classification is 0.93.

INDEX TERMS Breast cancer, classification, computer-aided diagnosis, deep learning, mammogram, multi-view, feature fusion, transfer learning.

I. INTRODUCTION

Breast cancer is the second most leading cause of cancer related deaths among women. Statistically, 266,120 new breast cancer cases were diagnosed in the US in 2018, which is the highest occurrence level of 30% among women [1]. World Health Organization (WHO) estimated that in 2018, breast cancer-related death exceeded 627,000 cases around the world [2]. According to the American Cancer Society (ACS), only in America, the average risk of a woman developing breast cancer over the course of her lifetime is about 12%, that is every eighth woman [3]. Studies continue to uncover lifestyle factors and habits such as feeding, smoking, drinking, and stress which may cause the disease [4].

Breast cancer is usually detected in its later stages. Early detection of breast cancer significantly increases the chances of survival. Mammography is a well-known and very effective method for breast cancer screening. A lot of differences (such as shape, size, distribution and boundary etc.) present

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

in breast lesions increase the misdiagnosis ratio [5]. Mammograms are normally taken earlier than biopsies. The latter is only done to confirm the findings of the former. Earlier detection can save lives, and mammograms are the first line of action when it comes to cancer diagnosis [6]. Hence, this effort is directed towards the use of mammography for early diagnosis of the disease. Much effort has already gone into diagnosis and detection of breast cancer through mammograms. Like other images, they are subjected to similar steps and procedures, such as segmentation, feature extraction, and classification.

Manual evaluation of mammograms is time consuming and is prone to errors. CADx system is a well-researched technique for improving the accuracy of mammogram based diagnosis and reducing the chances of error due to variation in human visual perception. This system helps the radiologists by diagnosing medical images more accurately and at low cost [7]. Computer aided detection and diagnostic systems are employed to act as the second opinion on medical images in many hospitals [8]. In traditional CADx systems, the focus has always been on handcrafted features learning techniques. However, over the last five years deep learning paradigm has been increasingly employed in computer vision related tasks. Deep learning has been commonly used in activities related to medical imaging [9]–[12]. It is based on the workings of the human brain and it models them in deep architectures to detect abnormalities in images.

Multi-view representation learning has been gaining significant interest in the artificial intelligence field in recent years. This work deals with the limitations of the current research in multi-view based CADx systems using deep learning. Previously single-view information from mammograms has been used for breast cancer diagnosis. In multi-view CADx systems, radiologists normally use four views information of mammograms to analyze the breast cancer prediction. Screening mammogram mostly provides four x-ray images of both breasts or two different views of each breast. These two views are CC (the view from above) and MLO is an oblique or angled view (taken under 45 degree). They start the analysis by looking at Left Cranio-Caudal (L-CC) view. If they find any abnormality in this view, then they prefer to check Left Medio-Lateral-Oblique (L-MLO), Right Cranio-Caudal (R-CC) and Right Medio-Lateral Oblique (R-MLO) views of the mammogram, respectively. The chances of abnormality increase with existence of suspicious regions in these views [13]. However, combining the information of multiple views of a mammogram increases the performance of the CADx system.

Most of the studies in multi-view CADx have normally focused on just two views of mammograms (CC and MLO) [14], [15]. Li *et al.* [13] drew the focus on using four-view mammogram based CADx systems. They point out the importance of its usage in modern CADx system to improve the accuracy of breast cancer prediction.

The main objective of this paper is to evaluate deep Convolutional Neural Network (CNN) to classify mammogram

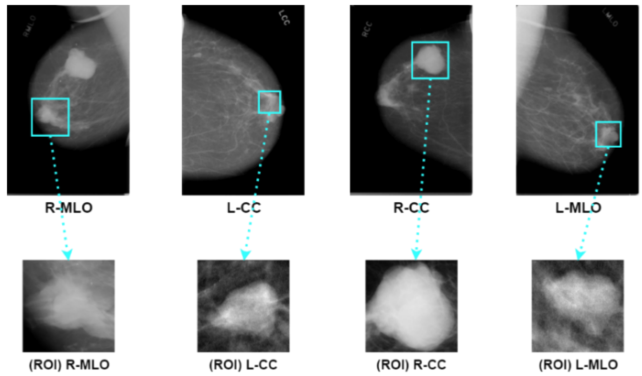


FIGURE 1. Mammograms and extracted patches of different views (Right MLO, Left CC, Right CC, and Left MLO) from CBIS-DDSM dataset.

into (abnormal or nomral), (mass or calcification) and finally in last (malignant or benign) using a fusion of four-views (R-CC, R-MLO, L-CC, and L-MLO). We can get useful feature information from all views in different training stages. Combining the data information from all the views before training is called early fusion and can be applied by concatenation of all features [16], [17]. The proposed framework based on the four-view feature fusion technique outperforms state-of-the-art approaches that are evaluated on benchmark datasets: CBIS-DDSM and mini-MIAS. Figure 1 shows the mammograms and extracted patches of different views (Right MLO, Left CC, Right CC and Left MLO) from CBIS-DDSM dataset.

In our assessment the main contributions of this paper are:

- We proposed a novel MVFF CADx system that is based on the feature fusion strategy of four views of mammogram of each patient which significantly improves the performance of classification.
- This paper presents a complete MVFF CADx tool which have the ability to classify any mammogram into normal, malignant and benign. Further, the tool can classify the mammogram into mass or calcification.
- In the framework of CADx of the breast cancer, we analyze the performance of augmentation on state-of-the-art CNN architectures through transfer learning between different domains (natural images and mammograms).
- The proposed MVFF CADx system not only boosts the performance of mammogram classification but also decrease the computational complexity. The reason behind it, we used simple network with less number of parameters instead of using complex networks which normally have large number of parameters.

This paper is organized into four sections. In section II, related work is presented. In section III, the methodology regarding this study is discussed. It provides details about data preprocessing, data augmentation, different CNN architectures applied and the Proposed Multi-view Features Fusion (MVFF) based CADx System. In section IV results are discussed. Our conclusions and future directions are presented in the section V.

II. LITERATURE REVIEW

Over the last few decades, significant work has been done regarding breast cancer diagnosis [18]–[21]. Levy et al. developed a transfer learning and data augmentation-based method to classify ROIs taken from DDSM dataset. Some proven pre-trained models such as AlexNet or GoogLeNet were used in this work. Their model achieved 92.9% accuracy by initializing the weights of pre-trained models [18]. Gardezi et al. considered abnormality classification in mammograms using deep learning features. They applied VGG16 CNN deep learning architecture on the IRMA dataset. The author faced the overfitting problem due to a smaller set of available images [19]. Charkraborty et al. proposed a novel system to detect masses in mammogram using high-to-low intensity thresholding and then multi-resolution analysis of the orientation of tissue is performed for classification of masses. They achieved 85% sensitivity for mass detection and AUC of 0.92 for masses diagnosis [22]. A CADx system for mass classification into malignant and benign was proposed by Zheng et al. [20] using a deep convolutional neural network. They claimed 0.8 area under the curve (AUC) performance of ROC. Significant work has been done regarding mass detection using state-of-the-art methods such as R-CNN and random forests [21]. All the above CADx systems were trained on just a single view however, these studies ignored the multi-view information of mammograms. Carneir et al. proposed an automated technique for two view mammogram (CC and MLO) analysis. Their system estimated the risk of breast cancer among women by observing different views of a mammogram. Their approach classified the whole mammographic image into micro-calcification and mass [15].

Gu *et al.* [23] proposed a multi-view-based auto-diagnostic system to improve the accuracy of CADx systems using CNN. Multi-view metric learning with global consistency and local smoothness (MVML-GL) applied in this work to view mammogram from a different view. Performance of risk assessment is increased using multi-view mammograms and segmented breast lesion. Their results showed that multi-view-based systems perform better than single view systems. Dhungel *et al.* [24] created and trained a multi-view deep residual neural network (mResNet) which consists of ResNet blocks. This pre-trained model has six input images including CC and MLO views as well as segmentation of breast lesion with multi-view. Their results showed that the proposed mResNet based on fully automatic classification produced reasonable results. Bekker *et al.* [25] developed a multi-view logistic regression-based classifier for microcalcifications indication into malignant or benign. They concluded the advantage of multi-view classifier by evaluating on standardized DDSM dataset.

Several studies other than medical imaging, for instance [16], [26], [27] also have been carried out on multi-view classification machine learning methods. Houthuys *et al.* [16] proposed a classification approach named as multi-view least squares support vector machines. Their main contribution is a coupling term for reducing the combination of error from multiple views. Few researchers have proposed four views-based systems [28], [29]. Wei *et al.* [28] proposed a CADx system which fused the extracted information of four views with a decision tree. The authors selected AdaBoost algorithm for training of decision tree after the comparison with linear discriminant analysis and support vector machine. Tan *et al.* [29] developed a breast cancer risk prediction model which is based on the three artificial neural networks. The model used two bilateral views of mammogram to fuse characteristics of images. They successfully demonstrated the effectiveness of a fusion of four-view mammographic features image which helped to predict the near term risk of breast cancer.

III. METHODOLOGY

In this section, we describe the data pre-processing steps used in this study, the data selection criteria, the data augmentation approach, most commonly used CNN architectures for the classification task, and the proposed strategy of four-view mammogram-based system.

A. DATA PRE-PROCESSING

Classification of raw medical mammograms without pre-processing steps do not create good results. We performed some well-known pre-processing steps like rescaling, bilateral filtering, contrast enhancement, etc., on our dataset for achieving better performance. Some quality distortion techniques were also performed.

B. DATA AUGMENTATION

In order to apply deep learning, a large number of images are required for proper training. A model trained on a smaller dataset tends to overfit. To minimize the effect of overfitting, we have augmented our training part of dataset more than seven times. The main variations adopted for image augmentation were random rotation up to 90 degrees, random vertical and horizontal shifts with a range of 0.2, random zoom with a range of 0.2, random shear with a range of 0.2 and nearest fill model.

C. CNN ARCHITECTURES

In this work, we evaluated four specific types of CNN architectures for training and fine-tuning strategies. In recent years, all of these well-known architectures have been the winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification benchmark [30]. We adapted these architectures for classification of mammograms on different levels (i.e.; normal vs abnormal, mass vs calcification and malignant vs benign) and compared their performance using four-views feature fusion strategy.

1) VGGNet

The main idea behind VGGNet architecture is much deeper networks with much smaller filters. VGGNet was introduced by Simonyan and Zisserman [31] from VGG (Visual

FIGURE 2. The basic architecture of VGGNet used in ILSVRC 2014 challenge. Different versions of VGGNet architecture have same basic structure, however the number of convolution layers changed in each architecture. Convolution layer is represented with 'Conv' and fully connected or dense layer with 'FC'.

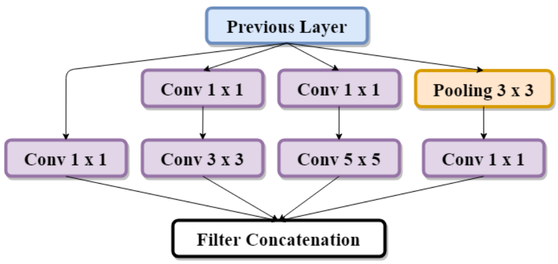


FIGURE 3. The basic Inception Module with dimensionality reduction of GoogleNet used in ILSVRC 2014 challenge. The bottleneck layers 'Conv 1×1' are convolutional layer of filter size 1×1 which was used for dimensionality reduction in the architecture.

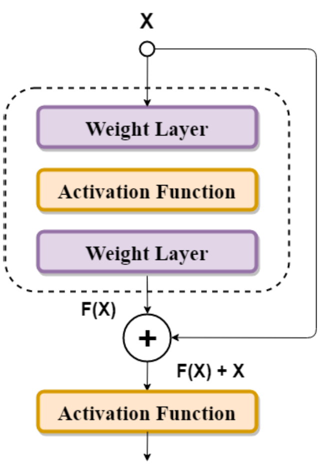


FIGURE 4. Residual Block of ResNet used in ILSVRC 2015 challenge. The ResNet architecture is 8 times more deeper than VGGNet but it has lower computational complexity than VGGNet.

Geometry Group), University of Oxford. This network performed very well in the ILSVRC 2014 challenge [30]. They performed best in the image localization task and were runners up in the image classification task with 7.3% top five error rate. The number of layers was increased from 8 in AlexNet [32] to 16 or 19, depending on the model, in VGGNet with a small 3x3 convolutional filter size. The two architectures with different layers were named VGG16 and VGG19 for obvious reasons. The 19 layers network performed slightly better but consumed more memory. Figure 2 show the architecture of VGGNet used in ILSVRC 2014 challenge.

2) GoogLeNet

In 2014, Szegedy *et al.* [33] from Google designed a much deeper network with 22 layers. GoogLeNet was the winner of the ILSVRC 2014 challenge in image classification task with 6.7% top five error rate. The special thing about this architecture is inception modules that make it very efficient in terms of computational cost. This architecture was able to save a lot of parameters due to the absence of fully connected layers. In this architecture, several types of filter operations were used in parallel rather than in a sequential manner. The feature depth of GoogLeNet was reduced using 1x1 convolutional bottleneck layers, which helped to manage computational complexity. Figure 3 shows the Inception module of GoogleNet used in ILSVRC 2014 challenge.

3) ResNet

In 2015, the residual neural network won the competition of ILSVRC with 3.6% top five error rate [34]. ResNet outperformed all previous architectures in almost all tracks i.e.; image classification, detection, and localization. It was a much deeper architecture with 152 layers. The ultra-deep architecture used a stack of residual blocks and each block had two 3×3 convolutional layers. The concept of residual mapping resolved the optimization problem in deeper models. The skip connections or residual blocks in ResNet helps the architecture to avoid gradient diminishing problem [35]. Figure 4 shows the residual blocks of ResNet used in ILSVRC 2015 challenge.

D. PROPOSED FOUR-VIEW FEATURES FUSION BASED CADx SYSTEM

It is a deep convolutional neural network model that has been used for training on four mammogram views of each patient separately. CBIS-DDSM (Curated Breast Imaging Subset of DDSM) and mini-MIAS datasets have been used for this system. Figure 1 shows a four-view sample of mammograms and their extracted augmented ROIs. The proposed CNN model extracts multiple features from four representations of mammograms. In the training phase, for getting the information from multiple views, early fusion is applied. Early fusion is the technique of merging multiple feature vectors into single feature vector [36]. The four-views' features' fusion based CADx system comprises of two strategies.

Extensive experiments with different variations in hyperparameters is performed before consideration of such network. We performed following variations of hyperparameters to choose best classification network:

- For the first stage of feature extraction, we changed the number of convolution layers between 3 to 7 layers. The experimental evaluation is also performed with max pooling and average pooling after all convolution layers.
- Two activation function i.e.; ReLU and Leaky ReLU are evaluated with convolution layers and last two fully connected layers. The sigmoid and softmax function are tested in final classification layer.
- Stochastic gradient descent with different initial learning rate from 1e-2 to 1e-5 is tested as optimization algorithm. The value of momentum is tuned between 0.7 to 0.9.

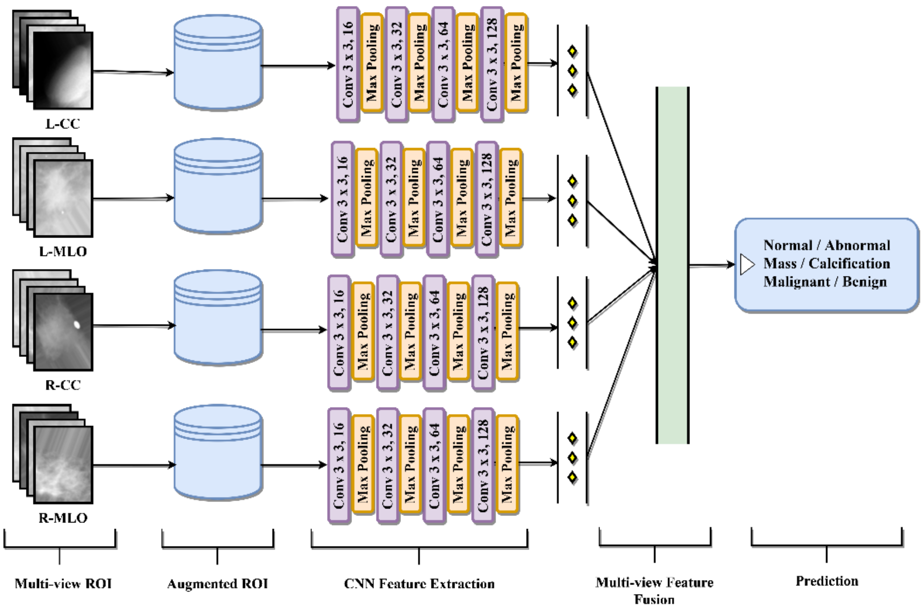


FIGURE 5. Proposed Multi-view Features Fusion (MVFF) based CADx system.

- The dropout after different convolution layers and full connected layers are evaluated with the ratio exist between 15% to 50%.

Under this proposed explaination, we have done several experiments and consider a network that outperformed other evaluated networks. The complete details about our proposed network is discussed in the next section.

1) FEATURE EXTRACTION STRATEGY

The first stage applies for feature extraction of each class of mammogram using deep learning. We have experimented with a different type of architecture in single or multi-view techniques. However, due to the simple nature of VGGNet, it performs better than others on the validation set for mammogram classification. We have done extensive experiments with several versions of VGGNet-like architectures for best architecture selection. The experiments showed that the architecture with (16, 32, 64, 128) layers performed best for the feature extraction stage. So, we decided to use this small version of VGGNet architecture for feature extraction using four view input of mammograms, as is shown in Figure 6. As in VGGNet, it is based on the 3x3 smaller convolutional filter size idea. The stride size of 1 pixel is fixed for each layer. A total number of four convolutional layers have been used in this network. The input image size fed to the network was fixed to 128 $\times$ 128. The maximum pooling with 2x2 kernel size was used for down sampling in image representation after each convolution layer. The activation function 'ReLu' was used with convolution layers. In this feature extraction step, the same convolutional neural network structure was used for each view separately.

FIGURE 6. Small VGGNet-like architecture used for feature extraction stage.

2) MULTI-VIEW FEATURE FUSION STRATEGY

The second stage is about the fusion of multiple features from various views. These features have been extracted from four views of mammograms in the first step. Mammography provides total of four views of mammogram (L-CC, L-MLO, RMLO, and R-CC) of each patient. There are two views of each breast from a different angle. In this early fusion technique, we concatenated all feature vectors of four views of mammogram into a single vector which helped to provide final classification. Our proposed system model is shown in Figure 5 and the complete MVFF network is summarized in Table 1.

IV. RESULTS AND DISCUSSION

In this work, we used the feature fusion strategy of four views to build a mammogram classification model. We also investigated the performance of state-of-the-art CNN architectures on the mammogram classification problem in order to give a comparison with our proposed multi-view model.

A. DATASETS

In this work, we aim to train and evaluate our model on two publicly available datasets. A final dataset was prepared by

TABLE 1. Structure of the Proposed MVFF Network.

Layer	Shape	Activation	Parameters
Input (1)	128x128x1	-	0
Input (2)	128x128x1	-	0
Input (3)	128x128x1	-	0
Input (4)	128x128x1	-	0
Conv2D (1)	3x3x16	relu	160
Conv2D (2)	3x3x16	relu	160
Conv2D (3)	3x3x16	relu	160
Conv2D (4)	3x3x16	relu	160
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Conv2D (5)	3x3x32	relu	4640
Conv2D (6)	3x3x32	relu	4640
Conv2D (7)	3x3x32	relu	4640
Conv2D (8)	3x3x32	relu	4640
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Conv2D (9)	3x3x64	relu	18496
Conv2D (10)	3x3x64	relu	18496
Conv2D (11)	3x3x64	relu	18496
Conv2D (12)	3x3x64	relu	18496
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Conv2D (13)	3x3x128	relu + dropout	73856
Conv2D (14)	3x3x128	relu + dropout	73856
Conv2D (15)	3x3x128	relu + dropout	73856
Conv2D (16)	3x3x128	relu + dropout	73856
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Max Pooling	2x2	-	0
Flatten-1	4608	-	0
Flatten-2	4608	-	0
Flatten-3	4608	-	0
Flatten-4	4608	-	0
Concatenate	18432	-	0
Fully Connected (1)	300	relu + dropout	5529900
Fully Connected (2)	300	relu	90300
Fully Connected (3)	2	sigmoid	301

combining both of these datasets in the context of normal vs abnormal, mass vs calcification and malignant vs benign classification in the breast region.

1) CBIS-DDSM

Digital Database for Screening Mammography (DDSM) is one of the most commonly used and largest publicly available datasets for the mammographic classification task. This dataset contains scanned mammography films of approximately 2620 studies. Each study includes four mammograms (i.e.; two images from each breast) with verified information of image and patient. An updated version of DDSM is Curated Breast Imaging Subset of DDSM (CBIS-DDSM), which is used for performance evaluation of our work. This subset of images from the original DDSM dataset is selected and curated by expert radiologists. These images have been compressed and converted into standardized DICOM format. The dataset contains two views of each breast (i.e.; CC and MLO), abnormality type and pathologic diagnosis. Table 2 shows the standardized train/test split of abnormality type.

TABLE 2. Standardized train/test split of abnormality type in CBIS-DDSM.

Abnormality Type	Training	Testing	Total
Mass	1318	378	1696
Calcification	1546	326	1872
Total	2864	704	3568

2) MIAS

The Mammographic Imaging Analysis Society (MIAS) dataset is another well-curated publicly available mammographic dataset. This dataset consists of 322 digitized films with radiologist's annotation on the position of abnormalities. All the images in this dataset have a resolution of 1024×1024 and contain left and right mammogram of each patient. Class and severity of abnormality are present in dataset details.

In all the experiments, the ROI size for input patches was 128×128 pixels which passed through the CNN models. We used following parameters for the training of deep learning models: the optimization algorithm used is stochastic gradient descent with initial learning rate as 1e-4, momentum was equal to 0.9, the batch size was 32, and categorical cross entropy as a loss fuction was used. The performance of trained models to correctly classify mammograms were evaluated on the validation set (20%) in a random fashion. The stopping criteria for training was set to 100 epochs and early stopping with the patience of 5 was also used. In this early stopping setup, the training process is terminated when the performance of network did not improve for five consecutive epochs or finished the cycle of maximum value of 100 epochs. The mammogram classification models were trained on NVIDIA Tesla P100, with 13 gigabytes of memory using keras (2.2.4) version and tensorflow (1.13.1) version at backend.

For transfer learning, all the deep learning architectures were pre-trained using ImageNet [30] dataset which contains 1000 classes. We fine-tuned these architectures for the two class problem by replacing the last three layers with two fully connected layers with 300 nodes each and a single classification layer. The number of trainable parameters and

TABLE 3. Parameters used for training of CNN models.

CNN Model	Trainable Parameters	Batch Size
VGG16	14,602,112	32
VGG19	19,911,808	32
ResNet50	23,524,992	32
GoogleNet	21,758,240	32
Proposed MVFF	6,009,109	16

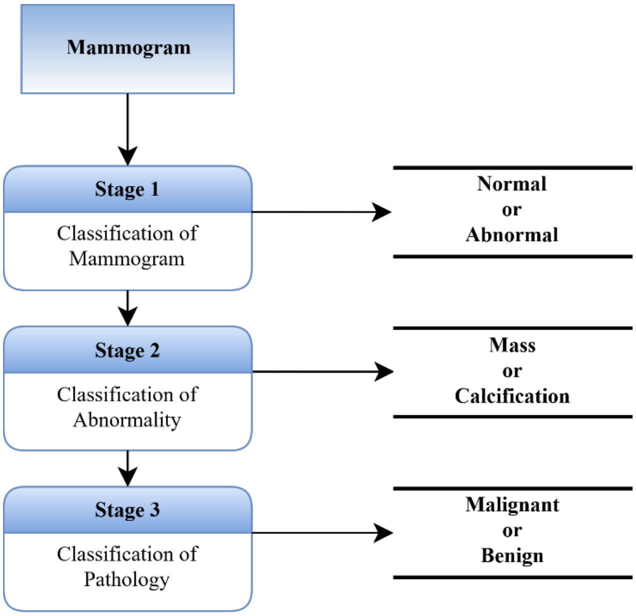


FIGURE 7. The structure of proposed complete CADx Tool for Mammogram classification in three stages.

batch size used during training of CNN models are shown in Table 3.

B. RESULTS OF CADx SYSTEM

The basic structure of our proposed complete MVFF CADx system is divided into three stages. In the first stages MVFF CADx system classify mammogram into normal or abnormal, in the second stage classification of mass or calcification performs and finally mammogram is classified into malignant or benign. A flow diagram of complete MVFF CADx system is shown in Figure 7.

The performance of CADx system to correctly classify mammogram for all stages are evaluated using standard medical metrics such as sensitivity, specificity, and accuracy. Sensitivity means that model predict the positive class of test mammogram when actual class is also positive (true positive rate). Specificity means that model predicts the negative class of test mammogram when mammogram also belong to the negative class (true negative rate). We also evaluated our models using ROC curve and area under the ROC for detailed analysis of proposed CADx system. The ROC plot is in two dimensional space where x-axis shows the false positive rate (1-specificity) and y-axis shows the true positive rate (sensitivity) of system. The strategy of five independent model

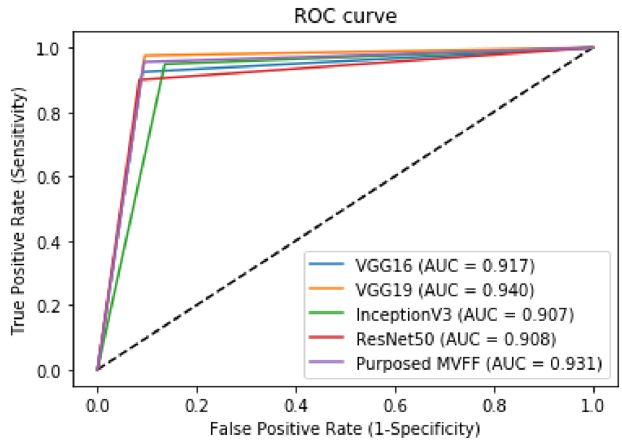


FIGURE 8. ROC curve for breast cancer classification into normal and abnormal on CBIS-DDSM: Testing performance of VGG16, VGG19, InceptionV3, ResNet50 and the proposed MVFF is plotted.

accuracies with mean and standard deviation was also used to evaluate classification performance. In the equation [1]–[3]: TP (True Positive), FP (False Positive), True Negative (TN), and FN (False Negative).

$$Sensitivity = \frac{TP}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

$$Specificity = \frac{TN}{\mathrm{TN} + \mathrm{FP}} \tag{2}$$

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{3}$$

a: STAGE 1: CLASSIFICATION OF MAMMOGRAM

In the first series of experiments, the classification of normal and abnormal mammogram was performed. The CBIS-DDSM dataset only contains abnormal images i.e.; mass and calcification images (see Table 2), and normal images are not available. However, for this stage of CADx system normal images are required. We used the mini-MIAS dataset to solve this issue. We randomly extracted the 128 x 128 pixel of patches from normal images of this dataset. A balanced dataset was prepared by randomaly selecting the equal number of patches from both classes. The results of experiments are listed in Table 4 which shows that VGG19 obtains the best result for mammogram classification with AUC of 0.940, a sensitivity of 98.07%, a specificity of 88.13% and accurcy of 93.45%. The proposed MVFF model was the second best model for this stage with AUC of 0.934, a sensitivity of 96.31%, a specificity of 90.47% and accurcy of 93.73%. The ROC curve of all models for normal and abnomral classification stage is shown in Figure 8.

b: STAGE 2: CLASSIFICATION OF ABNORMALITY

In this second stage of CADx system, the classification of abnormality type is performed. CBIS-DDSM dataset contains two type of abnormalities in mammogram: mass and calcification. According to the Table 5, we compared the results of some existing state-of-the-art deep learning architectures with our proposed MVFF architecture. The results

TABLE 4. Performance of CNN architectures using single view and proposed a multi-view architecture for normal and abnormal classification, in testing accuracy (μ and σ) represents mean and standard deviation for five independent training results.

CNN Model	Augmentation	Testing Accuracy (μ ± σ)	Training Accuracy (μ ± σ)	Sensitivity	Specificity	AUC
VGG16	No	91.57% ± 1.40%	92.33% ± 2.10%	94.30%	88.11%	0.917
	Yes	92.13% ± 1.92%	95.89% ± 2.71%	94.66%	88.91%	0.918
VGG19	No	93.45% ± 1.56%	97.92% ± 1.88%	98.07%	88.13%	0.940
	Yes	93.54% ± 1.29%	96.13% ± 2.29%	97.25%	89.09%	0.932
InceptionV3	No	89.98% ± 2.02%	94.94% ± 1.75%	95.88%	83.53%	0.907
	Yes	90.92% ± 1.95%	98.05% ± 2.82%	96.28%	84.92%	0.906
ResNet50	No	90.92% ± 1.81%	96.84% ± 2.04%	92.65%	88.60%	0.908
	Yes	91.67% ± 1.72%	94.42% ± 1.96%	93.44%	89.30%	0.914
Proposed MVFF	No	92.70% ± 1.33%	97.05% ± 2.63%	96.56%	88.09%	0.931
Architecture	Yes	93.73% ± 1.61%	96.66% ± 2.23%	96.31%	90.47%	0.934

TABLE 5. Performance of CNN architectures using single view and proposed a multi-view architecture for mass and calcification classification, in testing accuracy (μ and σ) represents mean and standard deviation for five independent training results.

CNN Model	Augmentation	Testing Accuracy (μ ± σ)	Training Accuracy (μ ± σ)	Sensitivity	Specificity	AUC
VGG16	No	89.48% ± 0.87%	93.26% ± 1.28%	90.16%	88.76%	0.897
	Yes	89.62% ± 0.68%	92.59% ± 1.72%	91.09%	88.14%	0.896
VGG19	No	86.26% ± 0.54%	90.43% ± 1.09%	89.74%	83.06%	0.864
	Yes	87.52% ± 0.94%	93.21% ± 1.56%	91.42%	84.00%	0.877
InceptionV3	No	86.82% ± 1.19%	88.12% ± 0.91%	86.67%	86.98%	0.868
	Yes	87.52% ± 1.36%	89.03% ± 1.23%	88.04%	86.96%	0.875
ResNet50	No	85.69% ± 0.73%	88,24% ± 1.68%	83.13%	89.03%	0.855
	Yes	85.55% ± 1.20%	90.90% ± 1.17%	82.28%	90.03%	0.862
Proposed MVFF	No	90.60% ± 0.92%	92.17% ± 1.35%	92.42%	88.80%	0.907
Architecture	Yes	92.29% ± 1.15%	95.68% ± 1.33%	93.37%	91.17%	0.923

TABLE 6. Performance of CNN architectures using single view and proposed a multi-view architecture for malignant and benign classification, in testing accuracy (μ and σ) represents mean and standard deviation for five independent training results.

CNN Model	Augmentation	Testing Accuracy (μ ± σ)	Training Accuracy (μ ± σ)	Sensitivity	Specificity	AUC
VGG16	No	73.47% ± 0.32%	80.51% ± 1.51%	76.42%	68.78%	0.721
	Yes	75.39% ± 0.98%	79.21% ± 1.72%	78.22%	70.98%	0.746
VGG19	No	69.11% ± 0.67%	75.27% ± 1.63%	71.93%	64.08%	0.671
	Yes	71.20% ± 0.55%	73.33% ± 1.34%	73.70%	66.83%	0.703
InceptionV3	No	68.24% ± 0.82%	76.12% ± 0.91%	71.43%	62.68%	0.663
	Yes	70.16% ± 1.12%	75.99% ± 1.23%	72.88%	65.38%	0.691
ResNet50	No	68.76% ± 0.25%	74.56% ± 1.68%	74.46%	61.29%	0.681
	Yes	69.98% ± 0.65%	74.08% ± 1.17%	75.46%	62.75%	0.691
Proposed MVFF	No	76.27% ± 0.89%	78.92% ± 1.35%	80.61%	70.37%	0.757
Architecture	Yes	77.66% ± 0.72%	80.56% ± 1.33%	81.82%	72.02%	0.769

of experiments shows that the proposed MVFF model outperformed the best performing mammogram classification models by a significant margin. The highest overall accuracy achieved by our proposed classification model is 92.29%, which is approximately 3% better than the second best model. The sensitivity of the model is 93.37%, specificity is 91.17%, the AUC is 0.923 and the ROC curve is shown in Figure 9. The worst case is obtained by ResNet50 with AUC value of 0.855, a sensitivity of 83.13%, a specificity of 89.03% and accurcy of 85.69%.

c: STAGE 3: CLASSIFICATION OF PATHOLOGY

The final stage of CADx system, a series of experiments for malignant and benign classification is performed. Table 6 summarizes the performance of pathology classification of multiple CNN architectures with a single view and proposed multi-view feature fusion CNN architecture. The proposed MVFF model achieved best AUC of 0.757, a sensitivity of 81.82%, a specificity of 72.02% and accuracy of 77.66%. Figure 10 plots the ROC curve for malignant and benign classification that shows the tradeoff

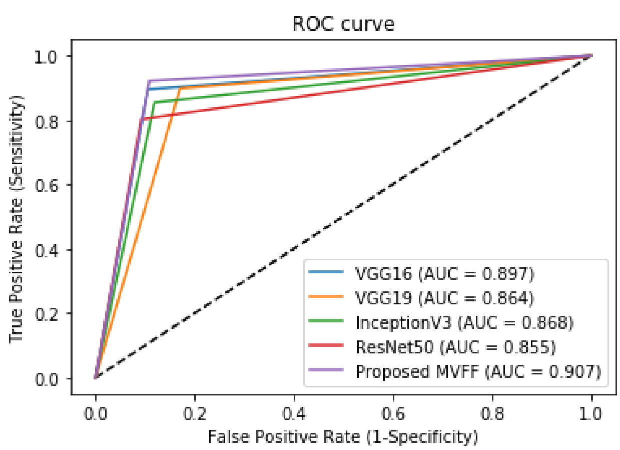


FIGURE 9. ROC Curve for breast cancer classification into mass and calcification on CBIS-DDSM: Testing performance of VGG16, VGG19, InceptionV3, ResNet50 and the proposed MVFF is plotted.

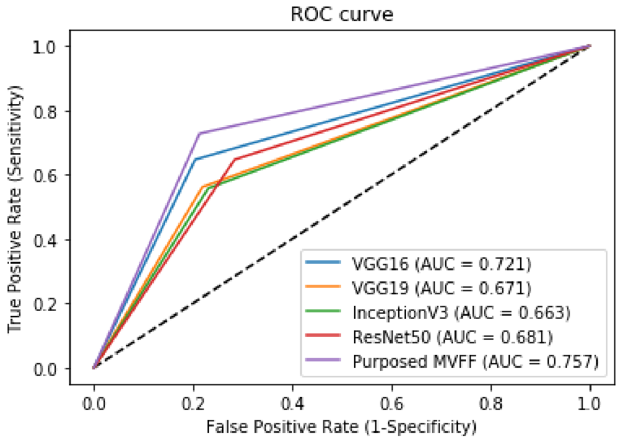


FIGURE 10. ROC Curve for breast cancer classification into malignant and benign on CBIS-DDSM: Testing performance of VGG16, VGG19, InceptionV3, ResNet50 and the proposed MVFF is plotted.

between true positive rate and false positive rate. The results in Table 6 demonstrates that our proposed MVFF model was able to surpass the performance of all other state-of-the-art model and proving the strength of our MVFF CADx system.

V. CONCLUSION AND FUTURE WORK

In this study, we proposed a MVFF model that utilizes the extracted information of four views of mammogram for classification. We concluded that multi-view feature fusion based CADx system is more efficient than single-view based system. The experimental results show the improvement of 2% to 4% in classification of mammogram using MVFF CADx for CBIS-DDSM and MIAS database, what confirms the effectiveness of this method. Furthermore, a comparison study depicts that small and simple architectures significantly perform better than the complex architectures in mammogram classification. According to our findings, augmentation plays an important role in MVFF based systems, because we faced overfitting problem in our experiments due to the unbalanced and limited number of images of each view. Data augmentation reduced the overfitting problem and helps to increase the testing accuracy from 3% to 5%. The strategy of data augmentation of mammogram with pre-trained weights of ImageNet did not achieve a big difference in mammogram classification accuracies. This strategy significantly performs better in complex architectures like GoogleNet and ResNet. We believe that our proposed MVFF CADx system introduces an important research topic in the domain of mammograms information fusion. We undertook this study to provide a baseline classification results of normal versus abnormal, mass versus calcification and malignant versus benign using MVFF strategy on CBIS-DDSM dataset.

In the future work, the developed methodology will be extended for the segmentation of mammogram. We will also focus on the use of the state-of-the-art architectures along with pre-trained weights for feature extraction stage and then use it into MVFF system.

ACKNOWLEDGMENT

This work has been carried out at Medical Imaging and Diagnostics Lab at COMSATS University Islamabad (CUI), Islamabad under the umbrella of National Center of Artificial Intelligence, Pakistan.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
[2] Breast Cancer, World Health Organization. Accessed: Nov. 15, 2019. [Online]. Available: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/
[3] American Cancer Society. *How Common is Breast Cancer?* Accessed: Nov. 15, 2019. [Online]. Available: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html
[4] K. McPherson, C. M. Steel, and J. M. Dixon, "ABC of breast diseases: Breast cancer-epidemiology, risk factors, and genetics," *Brit. Med. J.*, vol. 321, no. 7261, pp. 624–628, 2000.
[5] H. Li, S. Zhuang, D.-A. Li, J. Zhao, and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," *Biomed. Signal Process.*, vol. 51, pp. 347–354, May 2019.
[6] S. H. Heywang-Köbrunner, A. Hacker, and S. Sedlacek, "Advantages and disadvantages of mammography screening," *Breast Care*, vol. 6, no. 3, pp. 199–207, 2011.
[7] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. D. Heeten, and N. Karssemeijer, "Large scale deep learning for computer aided detection of mammographic lesions," *Med. Image Anal.*, vol. 35, pp. 303–312, Jan. 2017.
[8] P. Casti, A. Mencattini, M. Salmeri, A. Ancona, F. Mangeri, M. L. Pepe, and R. M. Rangayyan, "Contour-independent detection and classification of mammographic lesions," *Biomed. Signal Process. Control*, vol. 25, pp. 165–177, Mar. 2016.
[9] A. Hassan, M. Ghafoor, S. A. Tariq, T. Zia, and W. Ahmad, "High efficiency video coding (HEVC)–based surgical telementoring system using shallow convolutional neural network," *J. Digit. Imag.*, vol. 32, no. 6, pp. 1027–1043, 2019.
[10] E. Wu, B. Kong, X. Wang, J. Bai, Y. Lu, F. Gao, S. Zhang, K. Cao, Q. Song, S. Lyu, and Y. Yin, "Residual attention based network for hand bone age assessment," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1158–1161.
[11] H. Ye, "Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network," *Eur.radiology*, pp. 1–11, 2019.
[12] B. Kong, S. Sun, X. Wang, Q. Song, and S. Zhang, "Invasive cancer detection utilizing compressed convolutional neural network and transfer learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 156–164.
[13] Y. Li, H. Chen, L. Cao, and J. Ma, "A survey of computer-aided detection of breast cancer with mammography," *J. Health Med. Inform.*, vol. 4, no. 7, pp. 1–6, 2016.
[14] H. Wang, J. Feng, Z. Zhang, H. Su, L. Cui, H. He, and L. Liu, "Breast mass classification via deeply integrating the contextual information from multi-view data," *Pattern Recognit.*, vol. 80, pp. 42–52, Aug. 2018.
[15] G. Carneiro, J. Nascimento, and A. P. Bradley, "Automated analysis of unregistered multi-view mammograms with deep learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2355–2365, Nov. 2017.
[16] L. Houthuys, R. Langone, and J. A. K. Suykens, "Multi-view least squares support vector machines classification," *Neurocomputing*, vol. 282, pp. 78–88, Mar. 2018.
[17] 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.
[18] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," Dec. 2016, *arXiv:1612.00542*. [Online]. Available: https://arxiv.org/abs/1612.00542
[19] 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.
[20] X. Zhang, "Classification of mammographic masses by deep learning," in *Proc. 56th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Sep. 2017, pp. 793–796.

[21] 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)*, 2015, pp. 1–8.

[22] J. Chakraborty, A. Midya, and R. Rabidas, "Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns," *Expert Syst. Appl.*, vol. 99, pp. 168–179, Jun. 2018.

[23] X. Gu, Z. Shi, and J. Ma, "Multi-view learning for mammogram analysis: Auto-diagnosis models for breast cancer," in *Proc. IEEE Int. Conf. Smart Internet Things (SmartIoT)*, Aug. 2018, pp. 149–153.

[24] N. Dhungel, G. Carneiro, and A. P. Bradley, "Fully automated classification of mammograms using deep residual neural networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 310–314.

[25] A. J. Bekker, M. Shalhon, H. Greenspan, and J. Goldberger, "Multi-view probabilistic classification of breast microcalcifications," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 645–653, Feb. 2016.

[26] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.

[27] D. Tao, Y. Guo, B. Yu, J. Pang, and Z. Yu, "Deep multi-view feature learning for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2657–2666, Oct. 2018.

[28] J. Wei, H.-P. Chan, C. Zhou, Y.-T. Wu, B. Sahiner, L. M. Hadjiiski, M. A. Roubidoux, and M. A. Helvie, "Computer-aided detection of breast masses: Four-view strategy for screening mammography," *Med. Phys.*, vol. 38, no. 4, pp. 1867–1876, 2011.

[29] M. Tan, J. Pu, S. Cheng, H. Liu, and B. Zheng, "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk," *Ann. Biomed. Eng.*, vol. 43, no. 10, pp. 2416–2428, Oct. 2015.

[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[32] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

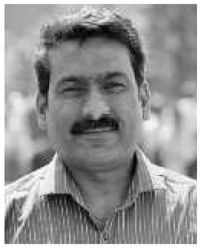
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[35] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Jan. 2019, *arXiv:1901.06032*. [Online]. Available: https://arxiv.org/abs/1901.06032

[36] A. Jouirou, A. Baâzaoui, and W. Barhoumi, "Multi-view information fusion in mammograms: A comprehensive overview," *Inf. Fusion*, vol. 52, pp. 308–321, Dec. 2019.

HASAN NASIR KHAN received the B.S. degree in computer science from COMSATS University Islamabad, Sahiwal, Pakistan, in 2016, where he is currently pursuing the M.S. degree. He is currently a Research Assistant with the Medical Imaging and Diagnostics Laboratory, COMSATS University Islamabad, under the umbrella of the National Center of Artificial Intelligence, Pakistan. He has five international conference proceedings. His research interest includes the development of computer-aided diagnosis systems for early diagnosis of breast cancer using artificial intelligence and computer vision techniques. He was a recipient of the Prime Minister of Pakistan's National ICT Scholarship Award, in 2012.

AHMAD RAZA SHAHID received the Ph.D. degree in computer science from York, U.K., in 2012. During his Ph.D., he worked on automatically building a WordNet for four languages, namely, English, German, French, and Greek. After his Ph.D., he has been working in the areas of computer vision and pattern recognition, machine learning, and natural language processing. A few of the problems that he has worked on include cancer detection, pedestrian detection, driver fatigue detection, and data mining. He is currently an Assistant Professor with COMSATS University Islamabad (CUI), Islamabad, Pakistan. He is also with the Department of Computer Science and Medical Image Diagnostic (MID) Laboratory, which is established under the National Center of Artificial Intelligence (NCAI).



BASIT RAZA received the Ph.D. degree in computer science from International Islamic University (IIU), Islamabad, Pakistan, in 2014. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan. He is also with the Department of Computer Science and Medical Image Diagnostic (MID) Laboratory, which is established under the National Center of Artificial Intelligence (NCAI). He has published a number of conference and journal articles of internal repute. His research interests are data science, data mining, image processing, machine learning, deep learning, and artificial intelligence.

AMIR HANIF DAR received the Ph.D. degree in computer applications from Beihang University, Beijing, China, in 2003. He is currently an Assistant Professor with COMSATS University Islamabad (CUI), Islamabad, Pakistan. He is also with the Department of Computer Science and Medical Image Diagnostic (MID) Laboratory, which is established under the National Center of Artificial Intelligence (NCAI). His current research is focused on medical diagnostic through different image modalities.

HANI ALQUHAYZ is an Assistant Professor in Computer Science department in College of Science at Majmaah University, Saudi Arabia. Hani has PhD in Computer Science from De Montfort University in UK. He got Masters in Information Systems Management and Bachelor in Computer Science from King Saud University. His research interests in Wireless security, Scheduling, image processing, IoT, Security and Privacy, Data Mining, Machine Learning and Artificial Intelligence. Dr. Hani has authored several papers in high impacted journals such as IEEE Access, *Sensors*, and *Wireless Communications and Mobile Computing*.