https://doi.org/10.24191/jeesr.v21i1.009

Improvement of Breast Density Classifier based on CNN Features Extraction and SVM in Mammogram Images

Noor Fadzilah Razali*, Iza Sazanita Isa, Siti Noraini Sulaiman, Noor Khairiah A. Karim, and Muhammad Khusairi Osman

Abstract— **Screening mammography has been clinically practiced as a common method for monitoring any potential breast diseases, especially in denser breasts among women. The advancement in medical imaging technology proved that the integration of artificial intelligence had given a significant impact on the breast screening process. However, due to various demographic patient backgrounds in clinical profile and non-standardized configurations of the developed intelligence models, such application is incapable of being applied by a health practitioner. Commonly, the deep learning method trained on a single network classifier has resulted in lower performance accuracy. With the motivation to improve the performance of classifying the dense breast, this paper proposes an improved classification model using deep learning approach of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) employed on different sets of publicly established mammogram images of craniocaudal (CC) and medio-lateral (MLO) views. In this study, pre-trained CNN models, namely GoogleNet, ResNet50, ResNet101 and AlexNet, are used with SVM as a classifier. The proposed method for the classification of breast density region is superior to the existing methods as indicated from model performance quantitative results of accuracy, precision and area under the curve (AUC) of its receiver operating characteristic (ROC) curves. Significant improvement in model performance has been obtained using ResNet50 and GoogleNet with SVM classifier with > 94% accuracy and AUC > 0.95. Furthermore, the model is proved to have good feature extraction capabilities to work for various breast density images that can be further explored into detecting malignancy from screening mammogram images.**

Index Terms— **Breast density, convolutional neural network, deep learning, support vector machine, mammogram, transfer learning.**

This manuscript is submitted on 22th December 2021 and accepted on 30th May 2022.

N. F. Razali[1*], I. S. Isa[1], S. N. Sulaiman[1,2], M. K. Osman[1] are with the [1]Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Permatang Pauh, 13500 Pulau Pinang, MALAYSIA and [2]Integrative Pharmacogenomics Institute (iPROMISE), UiTM Puncak Alam Campus, Bandar Puncak Alam, Puncak Alam, Selangor, 42300, Malaysia (e-mail: fadzilah708@uitm.edu.my, izasazanita@uitm.edu.my, sitinoraini@uitm.edu.my, khusairi@uitm.edu.my).

N.K.A.Karim, Regenerative Medicine Cluster, Advanced Medical and Dental Institute Universiti Sains Malaysia, Malaysia. (e-mail: drkhairiah@usm.my).

*Corresponding author
Email address: fadzilah708@uitm.edu.my

I. INTRODUCTION

BREAST-related diseases are often associated with the presence of malignant cells, which is prevalent in breast cancer cases, especially among women aged 40 and above [1]. In conjunction, screening mammography has become the gold standard as an initial screening method for breast cancer detection introduced in the 1970s [2] for early cancer detection to improve breast cancer mortality rate over the years [3]. Related to breast cancer occurrence, breast density is one of the risk factors physicians assess during mammogram screening and reporting. It is associated with a higher risk of having malignant lesions [4]. In addition, many worldwide countries have made it mandatory to report breast density levels in the patient's mammogram report, and the patients must be notified of any necessary further clinical actions [5]. Although high-density breasts tend to be seen but not limited to younger patients and have lower chances of developing malignant tumors [6], awareness among women motivate them to seek specialist services of finding out the breasts' health status through screening mammograms, especially those with a family history of cancer. As clinically practiced, an ultrasound (US) or 3D mammogram of Digital Breast Tomosynthesis (DBT) is commonly recommended by health professionals to complement screening mammograms for denser breasts since it obscure abnormalities within the fibroglandular breast region, which is visually difficult to make the clinical diagnosis [4], [5].

The Breast Imaging-Reporting and Data System (BI-RADS) introduced by the American College of Radiology (ACR) is commonly used for scoring [7] that categorized the lesion types such as masses and calcifications. Other attributes, including architectural distortion and breast density to associate final findings, were also concluded in a mammogram report [6]. The increased risk associated with dense breasts is approximately 1.6 to 2.2-fold, according to studies conducted in which single or multiple radiologists measured BI-RADS in a controlled manner [8] during the reporting process.

The reporting process is often interpreted differently that prone to biasing among different inter and intra-rater observers, including misdiagnosis. For example, in a nested case-controlled study of breast cancer patients, 75% showed the increased risk of developing cancer with cases that went unnoticed on screening mammography [9]. This is because the breast lesions visually have the same image intensity as the

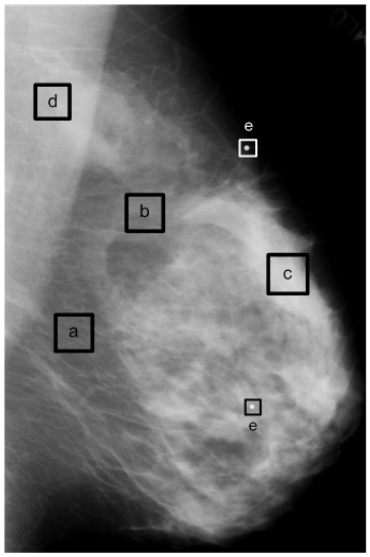

Fig. 1. Breast tissues within Level 3 density in MLO view image (a) Fatty region (b) Fibroglandular region (c) Dense region (d) Pectoral muscle (e) Possible artefacts.

surrounding fibroglandular tissues in the images were obscured by overlaying of morphologically similar surrounding fibrograndular tissues, where masking has been linked to lower detection rates in denser breasts [10]. Additionally, as shown in Figure 1, classifying breast density level is especially difficult among fibroglandular tissue density (b), and the presence of abnormality in the dense region (c) could also alter the judgment between inter-rater observers [11].

The growth of artificial intelligence (AI) in general has prompted researchers to investigate its application in medical imaging as a means of assisting tools in doing tedious and repetitive tasks to assist or at least at par with the function of an actual radiologist. To improve this, works on stages are made to create personalized breast assessments for different patients, such as automatically recognizing the prevalence of breast cancer gene-mutations and breast density level. Moreover, the development of computer-aided detection (CAD) system that implements deep learning as a subset of AI has been performed and integrated the algorithm in mammograms, especially in detecting cancer through various methods [11]–[15]. For example, a recent study conducted by [14] has employed mammogram images in their proposed model to detect cancer cases amongst UK and US patients. The result of the study has proved to reveal a lower false detection rate of human rater by 5.7% [14]. Nevertheless, breast density level is proven to be a major risk contributing to the presence of cancer through risk assessment studies by [16], [17] along with less developed density classification model developed according to specific demographics and limited manufacturer's images during training [7], [18].

Motivated by this, our goal is to propose a deep learning model that automatically learns features of density regions in mammograms and classifies them into non-dense and dense categories of breast mammograms. Finally, it will be tested on other manufacturer images. The features are extracted using several state-of-the-art pre-trained Convolutional Neural Network (CNN) models for comparison by employing the transfer learning method. The improvement of the model includes the implementation of the Support Vector Machine (SVM) as a classifier and is assessed through several model evaluation matrices. Furthermore, to assess the problem of generalizability caused by training using a single manufacturer's images, the algorithm's performance is tested on another manufacturer's dataset. In addition, a basic image pre-processing method is applied to the images prior to training to increase the system's performance further. The rest of the paper is organized as follows. Section II reviews some related literary works. In Section III, the details of our methodology are presented. Finally, section IV is the result and discussion, whereby Section V is the conclusion of this work.

II. Literature Review

Due to the prevalence of mammography as a key screening technique for breast cancer diagnosis, researchers in image processing have emphasized work on classification and detection algorithms employing these images. Common objectives among these researches are to improve false-positive diagnoses among radiologists, as well as to reduce the workload when done manually. The methods developed, including semi-automated breast density classification using the conventional method, using hand-crafted features extracted from each class to acquire unique characteristics as feature vectors extracted from information such as edges, intensity, and contrast from the mammogram training images [19], [20]. For example, studies conducted by [17] and [21] developed a semi-automated measuring approach for extracting features of intensity and textural patterns of gray-level co-occurrence matrix (GLCM) on the denser area. The odds ratio (OR) per standard deviation of the unadjusted cross-sectional measure was calculated using logistic regression and provided as the OR per standard deviation of the unadjusted cross-sectional measure with the additional info of the patients' body mass index (BMI) and age resulted in OR of 1.90 (95% confidence interval (CI) 1.73 to 2.09), equivalent to a fourfold interquartile risk ratio [6], [21]. Meanwhile, research done in [20] compared performances of using different feature vectors of the mammogram images through different kernel sizes of Law's mask analysis, trained using several machine learning classifiers such as probabilistic neural network (PNN), the neuro-fuzzy classifier (NFC) and SVM with the highest accuracy at 91.2%.

In practice, two machine learning-based tools accessed by radiologists to assess breast density falls into two assessment categories: volumetric such as Volpara, and area-based (i.e. Laboratory for Individualized Breast Radiodensity Assessment - LIBRA) methods [21]. The difference between the two methods relies on the inclusion of skin during scoring estimation. However, due to the higher positively skewed distribution of volumetric breast density, estimation obtained from volumetric algorithms are more difficult to correlate with radiologists' assessments [21], resulted in radiologists receiving significantly lower scores than expected. This shows that breast density categorization tools in two-dimensional aspects are beneficial for classification purposes [7], [22]. Although proven to have significantly high accuracy, semi-automated method

requires in depth knowledge of the input images to extract hand-crafted features associated with different objectives of the study with radiologist's guidance as well as multiple trials needed to ensure the specific descriptor as absolute to the algorithm [20], [23].

To overcome this, a newer deep-learning approach includes the use of the CNN model that bypassed the feature selection method by automatically learning all features in the image [24]. The implementation of a CNN-based algorithm enabled the development of a single end-to-end model that directly used the image with minimal pre-processing steps. However, through literature search, most deep-learning-based algorithm using breast mammogram studies for abnormality classification such as [12], [14], [19], [25]–[27] as opposed to classify breast density [7], [10], [11]. As example for density classification, [11] classify levels 2 and 3 of breast mammogram images by implementing AlexNet CNN using 14,000 normal mammogram images using scratch and transfer learning method that showed transfer learning was able to adapt to scratch based-CNN to a negative margin of only 0.0025% accuracy trained using balanced dataset. Due to the unavailability of ground truth in the dataset of this study, the authors made comparison by calculating percentage density (PD%) estimation using LIBRA software. In a study using FFDM INbreast public dataset by [18], the author pre-resized and augmented the images 4-fold of its original training number, and this resulted in increasing performance tremendously by implementing transfer learning on three pre-trained CNN models, with the highest accuracy of 99.72% using DenseNet to classify images of dense or non-dense based on the presence of a malignant lesion. Additionally, [10] applied transfer learning to detect malignancy on exclusively collected mammogram images from various breast density level images. The study concludes that with a higher density of breast, the CNN performance lowers to 84.3% detection accuracy at level 4 as the effect of masking of the lesion on dense region compared to level 1 at 95% detection accuracy [10]. Meanwhile, [28] combined levels 1 & 2 and levels 3 & 4 as dense and non-dense for classification using scratch-based CNN model and also employed similarity index for segmenting asymmetric region between both breasts that, emphasizing the importance of segmentation and image registration before detection and classification protocol in the algorithm. These studies show the classification task by the implementation of state-of-the-art CNN architectures into own dataset were able to lessen the computational burden by transferring its learned weights into own case through transfer learning method, with significantly high performance. However, using a single manufacturer for training is another issue faced by researchers to make sure their algorithm is applicable in clinical settings that include various manufacturers. Authors in [7] addressed this issue by combining three manufacturers' mammogram images in their InceptionV3 network to classify the images into lower risk (non-dense) and higher risk (dense) using cropped MLO images. The author catered the imbalanced dataset issue by tuning the class weight of the loss function for training. A variety of training input data were able to enhance the testing performance when tested on other manufacturer's images' to 92% testing accuracy.

Introducing SVM as a CNN-hybrid classifier substituting softmax layer has been applied that showed significant improvement in various studies, including breast abnormality [12], [19], [25], [26], to increase the performance of CNN based algorithm further. An earlier study by [19] applied CNN and SVM specifically for breast lesion detection through multiple pre-processing steps, using geostatistics function as shape descriptor features which were then used to classify masses region within an image. The features were used to segment suspected areas to be trained by the CNN and able to have 84.62% detection accuracy. A semi-automated study that implemented SVM to classify dense and non-dense mammogram images showed the highest performance of 91.6% using SVM compared to other classifiers such as PNN and NFC in [20], also by using texture features extracted manually. Meanwhile, both [12] and [25] applied SVM with CNN as the main training backbone to classify lesion abnormality in mammograms and show the accuracy of 97% using the MIAS dataset. Additionally, [25] suggested that SVM may offer better optimization and is nearly related to the dataset's diversity and non-linear nature. Moreover, the SVM was able to achieve an easier global minimum and withstand more noise, making it imperceptibly more resilient to a large number of parameters. As a result, SVM models may offer significant convergence and robustness advantages over standard CNN models. As breast densities are characterized by the variance of image intensities, as seen in Figure 1, this suggests a good implementation of SVM on classifying breast densities.

From the literature, it can be summarized that the methods to fully automate density classification based on intelligence approach is severely limited as its significance for its association with cancer risk together in clinical profiling is important in mammogram image reporting. This could help the radiologists to screen the risk of potential cancerous mass in the breast through an early detection approach. The major challenges of the accurate breast density classification using deep learning approach lie in the variety of interpretation of breast density levels, low signal-to-noise level in mammogram images, presence of lesions in training image, and lack of large datasets with density and abnormalities ground-truth available for researchers [11], [24]. Therefore features extraction method is a fundamental task in the CNN model to improve the correct classification and further facilitate radiologists in their work. In this paper, a classification model based on CNN features extraction and SVM classifier is presented to improve the system's performance in classifying breast density with the motivation to minimize the training losses of the model without lessening the efficacy.

III. Methodology

A. Mammogram Dataset Preparation

1) CBIS-DDSM

The Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is a publicly available mammogram dataset that is

the subset of its original database (Digital Database of Screening Mammogram - DDSM) comprising of 2620 digitized screen filmed mammograms obtained through the collaboration of Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department in 1997 with multiple ranges of image resolutions [29]. It is a publicly available mammogram that incorporates breast lesion images ranging from masses and calcification, supplemented with verified reports for each case including pathological categories, density level, BI-RADS categories and malignancy types. This study includes images from masses for training corresponding to different density classes. In addition, both breast views of craniocaudal (CC) and medio-lateral oblique (MLO) are treated as separate image cases for training. Since it was formed as screen-filmed mammogram, the CBIS-DDSM dataset is considered as having low image quality compared to a full-field digital mammogram (FFDM). However, it is extensively used in previous studies, especially in studies for the classification of breast abnormality [12]–[15] and breast density classification [30]. For the purpose of binary classification of the dense and non-dense breast in this study, classes 1 and 2 from the training set are combined as 'Non-Dense' as depicted in Figure 2 (a) and (b), while the remaining 3 and 4 as 'Dense' images as shown in Figure 2 (c) and (d). To develop and evaluate the system's performance, 80% of the images are randomly used for training, while 10% is used for validation. The remaining 10% is reserved as a never-seen dataset for testing purposes.

2) MIAS

A second dataset is used to assess the trained model's capacity to be tested on different datasets. The Mammographic Image Analysis Society (MIAS) are mammogram images obtained of the same group name from the UK Breast Screening Program [31]. It contained 322 whole mammogram images of 1024 x 1024 resolution for normal, benign and malignant classes of MLO breast view only. The accompanying report includes ground-truth from radiologists, including the coordinates of suspected tumors location, breast tissue types and abnormality morphology. Like CBIS-DDSM, MIAS is consisted of a decompressed version of digitized screen-filmed mammogram that was then converted into portable gray-map image format during its post-processing stages [31]. The images categorized into non-dense and densely labeled as 'Fatty - F' and 'Dense - D' in its supplementary ground-truth report are chosen for the testing stage to analyze the developed system's ability to be used on other images aside from the training images.

B. Proposed Method

This section discusses the overall proposed methodology to develop the system. Figure 3 depicts the overall schematic representation flow of the proposed experimental setup into stages of image pre-processing, CNN training and feature extraction, classification, and finally, performance evaluation. This experiment is conducted on a workstation equipped with an Intel Core™ i7-10870H 2.3GHz processor, NVIDIA GeForce RTX2060 single graphic processor unit (GPU) graphic card, and 16 GB RAM memory for the purpose of training,

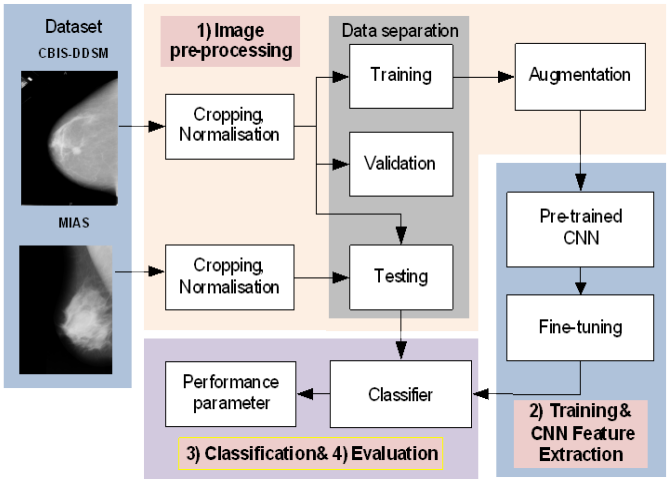

Fig. 3. Overall methodology for improvement of density level classification model on mammogram images using Convolutional Neural Network (CNN) and classifiers.

tuning and testing the network. The system is developed in MATLAB© (Release R2021b, Mathworks, Natick, MA).

1) Image Pre-processing

All the images used for training, validation and testing are pre-processed prior to being introduced in the system. As the dense region is sensitive to the intensity level of the image, fewer modifications, including image enhancement was not applied to the image to maintain its originality. All manual annotations overlayed on the images were digitally removed while extracting the largest blob in the image (hence, the breast region) through image processing techniques. As the presence of pectoral muscle might be disguised as a dense breast region, it is also removed from all MLO view images from both datasets. The background region of the images is automatically limited to only include the breast region to reduce the image dimension subsequently lower the computation load. All the images are normalized to ensure the data distribution of each input pixel is identical and accelerates convergence while training the network. Original image size is retained and resized later, corresponding to different input image size requirements of different pre-trained networks. Double rescaling prior to training might cause loss of pixel information in exchange for longer computation time. To reduce the effect of overfitting, augmentation techniques including random rotation and rescaling are applied to the training images.

2) Training & CNN feature Extraction

Four pre-trained networks are employed and tested on both datasets, namely as GoogleNet [32], AlexNet [33], ResNet50 and ResNet101 [34]. The overall proposed transfer learning process is illustrated in Figure 4. The training process includes the feature extraction layers and fine-tuning of the learned model. The feature extraction process in CNN involves the repetition of three basic CNN

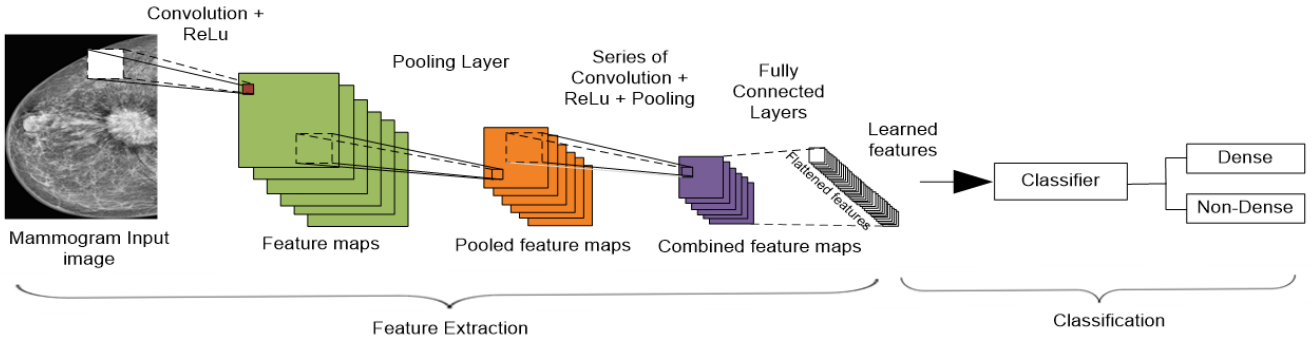

Fig. 4. Proposed transfer learning process with feature extraction and classification using the network's layer and Support Vector Machine (SVM) for breast density classification

layers, which are the convolutional, pooling layer and its activation function. The framework design of these three layers depends on the type of pre-trained network used. Rectified Linear Unit (ReLu) non-linear activation functions are used in all pre-trained CNN in this study. This activation function is preferred by the CNN developers because of its ability to overcome vanishing gradient problems that allow faster and better performance of the model [27].

Earlier CNN layers are used to extract visually interpreted high-level image features that is usually observed firsthand such as the edges and corners [32], [33]. In contrast, low-level features, for instance, the images' spatial and contrast information, are captured towards the end layers prior to entering the fully connected layer (FCL) [32], [33]. Finally, the extracted learned features are concatenated for the final flattened features before entering the classification stage. Fine-tuning is done by varying the training parameters such as the learning rate and choosing best mini-batches number per iteration to get the best result. The selected training parameters are set to default for ideal comparison between

TABLE II
TRAINING PARAMETERS FOR ALL PRE-TRAINED CNN MODELS.

Parameter	Settings
Optimizer	Adam
Initial Learning Weight	0.0001
L2 Regularization Rate	0.0001
Epoch	40
Mini-batch Size	30

all models, as summarized in Table II. This study utilized the network's own classifier by using the softmax layer right after the FCL layer and also by using SVM..

3) *Classification*

For classification stage, classifiers of each pre-trained network are compared to the SVM classifier to evaluate the performance ability of all the models to differentiate dense and non-dense regions in mammogram images.

a) Pre-trained CNN classifier

In a common CNN network, the classification layer consists of two layers: activation function (softmax), and the classification layer comes immediately after the FCL layer. A softmax function calculates the probability of each class from which the input belongs, where the probability, f of each class are defined as in Eq. 1.

$$f(y_i) = \frac{e^{y_i}}{\sum_k e^{y_k}}. \quad (1)$$

Once a test input image goes through the model for testing, the neural output of y is calculated for each feature class of the trained network. Then, the exponential of a particular class is computed over a total of the exponential of each class with the corresponding neural output to produce the probability value f. The highest value of f between the classes will be considered as the final class prediction.

b) Non-Linear Support Vector Machine (SVM)

In this study, a supervised learning algorithm SVM classification is applied for the comparison of performance between CNN and SVM classifiers. The original CNN layer is frozen after the FCL, and the resulted features are used as input to SVM. Determination of the hyperplane using the SVM algorithm functioned to separate the input features between the classes. The algorithm works by considering a feature to be classified represented as a point in n-dimensional space as the coordinate of this point (features). SVM performs the classification task by drawing a 2D or 3D hyperplane in such a way that all points in one category are at one side or the other side of the hyperplane. There could be more than one hyperplane that exists between the categories, and the SVM tries to separate the best way between the two categories, in a sense that it maximizes the distance (margin) to all points in either category. If a point falls into this margin, it is called the supporting vector, where the constraints of the points of each category should be on the correct side of the hyperplane. The algorithm works by solving a convex optimization problem that maximizes the margin γ, where:

$$\max_{w,b} \gamma(w,b), \quad (2)$$

in such that the linear separating hyperplane Eq. 3:

$$\forall i \; y_i(w^T x_i + b) \geq 0, \quad (3)$$

where (w,b) is a set of points of the optimal hyperplane in the linear line of x and y. However, the simplicity of the

SVM can be a problem if the points cannot be clearly separated by the hyperplane in larger training features [19]. Overcoming this problem includes increasing the training data for imbalanced class problems through augmentations, finding the separating hyperplane in the higher dimensional space and projecting back to the original space. In this study, the computation of the margin of the hyperplane is optimized as its input features were already well learned by the trained CNN network layers. The setting parameters for SVM are summarized in Table III.

TABLE III
SETTING PARAMETERS FOR NON-LINEAR SVM.

Parameter	Settings
Optimizer	Bayesian
Learner	Non-Linear
Coding	onevsall
ObservationIn	Column
Function Evaluation	30

To conclude, a newly proposed method for improving the performance of a pre-trained model to classify dense regions in mammogram images is proposed using CNN features extraction and SVM classifier as a learning algorithm. This improvement method for classifying the dense regions based on density level is summarized as follows:

1. The pre-processing of the mammogram images is conducted to remove unwanted regions such as pectoral muscle and manual annotations in the background.
2. Image normalization and image resizing are applied as attributed to the pre-trained CNN network.
3. The features information from the image is extracted using the feature extraction function of the pre-trained network.
4. Based on the default training parameters of the pre-trained network, the features extracted from the image is classified using the SVM algorithm as a learning classifier to replace the pre-trained CNN classifier.

4) Performance Evaluation

In this study, the model's performance parameters are calculated using the values from the confusion matrix, which are the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for each model. The accuracy, sensitivity, specificity, and precision are then computed as denoted by Eq. 4 to Eq. 7.

$$\text{Accuracy} = \frac{\text{TP+TN}}{TP+FN+TN+FP} \quad (4)$$

$$\text{Sensitivity/recall} = \frac{\text{TP}}{\text{(TP+FN)}} \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{(TP+FN)}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{(TP+FP)}} \quad (7)$$

Additionally, the system's performance is evaluated using its area under the curve (AUC) calculated from its Receiver Operating Characteristics (ROC) probability curve, which is calculated by plotting true-positive vs false-positive rates for each models. The improved model is named 'Pre-trained CNN + SVM', and its performance evaluation is per discussed in Tables V and VI. Gradient-weighted Class Activation Mapping (Grad-CAM) heat map is employed on sample images to visualize the critical localization of the image during testing on lower and higher performance models for qualitative comparison.

IV. RESULT AND DISCUSSION

TABLE IV
PRE-TRAINED MODELS SPECIFICATION AND TRAINING RESULTS

CNN Architecture	Para-meters (million)	Net-work Depth	Training Time (mins)	Final Training Loss	Final Vali-dation Loss	Vali-dation Accuracy (%)
GoogleNet	7	22	54.32	**0.34**	**0.42**	84.03
ResNet50	25.6	50	53.20	0.25	0.61	**92.09**
ResNet101	44.6	101	55.03*	0.27	0.52	78.60
AlexNet	61	8	57.93	0.39	0.55	83.33

*stopped early due to signs of overfitting at epoch 30

A. Result of pre-trained CNN models for training & validation

The training performance using CBIS-DDSM dataset for all four pre-trained CNNs and their respective specifications can be observed in Table IV, showing the highest validation accuracy using the ResNet50 model (92.09%).

However, it is shown to be prone to overfitting as the training progresses as the validation loss differs from its validation loss to 0.36 margin in the final epoch training. This suggests the lower mini-batch number could be employed during training in the future. All four networks were trained in significantly close training time margins with the same epochs and mini-batches per iteration except for ResNet101. An early stopping method is applied when signs of overfitting appear and performance plateaus during training. The best training vs validation loss was observed in the GoogleNet model, having a margin of ±0.08, as can be seen in Figure 5, where the best loss occurred at iteration 21x30 = 630th. The trend shows validation data represents the training dataset closely, although it appears noisy on some points, possibly due to the low number of validation data.

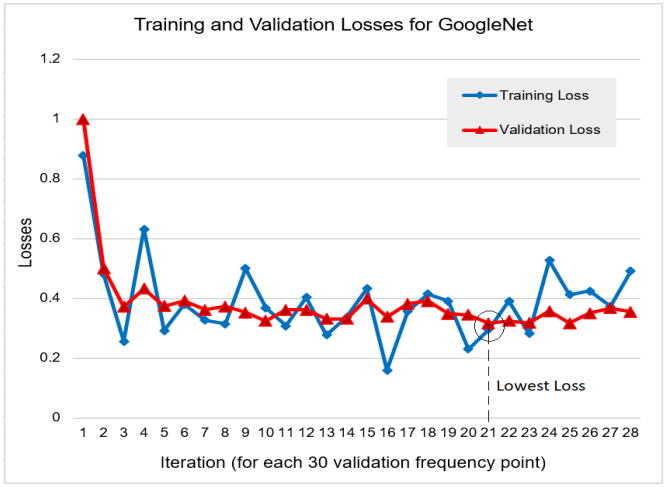

Fig. 5. Best Training and Validation Lossess observed on validation frequency points during training of GoogleNet model.

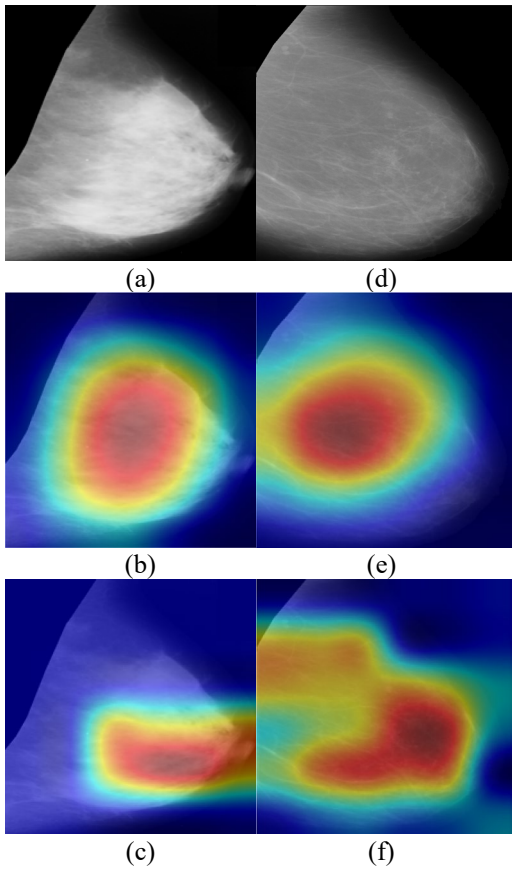

Fig. 6. Original (a) Dense and (d) non-dense sample breast image and their respective foci of network prediction on correctly classified image using Grad-CAM, on GoogleNet for (b) and (e), AlexNet for (c) and (f).

To qualitatively observe the system's performance after being trained, grad-CAM was applied with a sample of labeled test images, as shown in Figure 6. It showed the ability of the model to correctly identify the appropriate region to focus during prediction on the test image through Grad-CAM, where the red region emphasized dense (left) and non-dense (right) regions, respectively, on the correctly classified images. The comparison was made between two CNN models of GoogleNet (middle layer, Figure 6 (b) and (e)) and AlexNet (bottom layer, Figure 6 (c) and (f)) for both images. The red region on represents the region focusing on the density map of the images that are used to determine the class of the test images. Based on the second-best loss observed during training in Table IV for the AlexNet model (margin = ±0.16), the comparison is made to determine the best training model through Grad-CAM. The result shows scattered red region mapped, including out of breast region observed in AlexNet model heat map, compared to more focused red region observed from GoogleNet model. This observation significantly suggests better performance expected from GoogleNet compared to AlexNet models.

B. Performance Comparisons of Pre-trained CNN models on different Datasets for models using CNN classifier

Table V shows the result of the performance for each model tested on CBIS-DDSM and MIAS dataset, trained and classified using the original CNN pre-trained models. For CBIS-DDSM, the highest performance is observed using

TABLE V
BREAST DENSITY CLASSIFICATION USING PRE-TRAINED CNN MODELS FOR DIFFERENT TEST DATASET.

Dataset	Trainer / Classifier	Accu-racy	AUC	Sensiti-vity	Speci-ficity	Preci-sion
CBIS	GoogleNet	0.8112	0.8572	**0.7857**	0.8174	0.5116
	ResNet50	**0.8392**	**0.9009**	0.7273	**0.8889**	**0.7442**
	ResNet101	0.7972	0.8809	0.675	0.8447	0.6279
	AlexNet	0.7902	0.7819	0.7407	0.8017	0.4651
MIAS	**GoogleNet**	**0.9493**	**0.9941**	**1**	0.906	0.9009
	ResNet50	0.8783	0.9869	0.9869	**0.9902**	**0.9130**
	ResNet101	0.9217	0.9831	0.97	0.8803	0.8739
	AlexNet	0.8756	0.9857	1	0.797	0.7568

ResNet50 at 83.92% accuracy and AUC of 0.9009, along with the highest other metrics except for sensitivity scores. Meanwhile, when tested using another dataset MIAS, the result showed the highest accuracy using the GoogleNet model at 94.93% with an AUC of 0.9941. Moreover, the same image pre-processing were made to both datasets prior to the testing stage, significantly suggesting the models' generalizability and applicability to other datasets.

Overall, higher performance is seen when tested with the MIAS dataset for all models compared to the original dataset. This might be rooted in the variance of image density levels (4 ground truth density levels) that contributed to its lower performance than the dataset used from MIAS (3 ground truth density levels). However, it is worth noting that both CBIS-DDSM and MIAS are from different mammogram vendors where the difference lies in its data acquisition process, including radiation and contrast settings as well as post-processing methods of extracting images to different formats (DICOM and .pgm) might further alter the differences between the two mammogram datasets. To conclude, both datasets could be classified at acceptable performance, even without model improvement being applied in the result yet.

C. Result of Improved pre-trained CNN model using SVM classifier

The comparison results between the pre-trained CNN and SVM classifiers are tabulated in Table VI using the MIAS dataset. The same pattern can be observed in all models except for ResNet101, showing better overall performance when applying SVM as a classifier to the testing set. Best accuracy is observed in GoogleNet + SVM at 95.39% accuracy and highest AUC at 0.9949, as well as the highest sensitivity, specificity and precision scores. Meanwhile, ResNet101 + SVM has lower accuracy (91.71%) by 0.46% when classified using the original

TABLE VI
BREAST DENSITY CLASSIFICATION USING IMPROVED PRE-TRAINED CNN AND SVM CLASSSIFIER USING MIAS DATASET.

Trainer / Classifier	Accu-racy	AUC	Sensiti-vity	Specifi-city	Pre-cision
GoogleNet	0.9493	0.9941	1.0000	0.9060	0.9009
GoogleNet + SVM	**0.9539**	**0.9949**	**1.0000**	**0.9138**	**0.9099**
ResNet50	0.8783	0.9869	0.9869	0.9902	0.9130
ResNet50 + SVM	**0.9401**	**0.9869**	0.9900	0.8974	0.8919
ResNet101	**0.9217**	0.9831	0.9700	0.8803	0.8739
ResNet101 + SVM	0.9171	**0.9833**	0.9794	0.8667	0.8559
AlexNet	0.8756	**0.9857**	1.0000	0.7970	0.7568
AlexNet + SVM	**0.9078**	0.9850	0.9789	0.8525	0.8378

CNN softmax function (92.17%) but better AUC at 0.9833. A significant improvement trend is observed when SVM is applied for both datasets when other parameters are included on AUC, sensitivity and precision scores.

Figure 7 shows the models' accuracy on both CBIS-DDSM and MIAS datasets using original pre-trained CNN and SVM

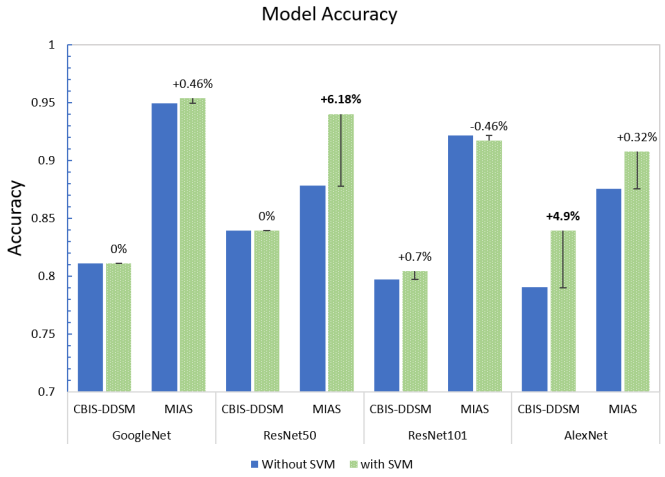

Fig. 7. Pre-trained CNN model accuracy based on SVM improvement on test datasets.

classifiers. From all 8 models trained in this study, 5 models were improved using SVM classifier with the highest accuracy upgrade by +6.18% can be observed when improving the model using SVM with ResNet50 in MIAS dataset, and increase of +4.9% in AlexNet for CBIS-DDSM test images.

From the results, the proposed improvement using SVM classifiers are proven to increase the system's performance in detecting dense and non-dense regions. Furthermore, non-linear kernels applied during hyperplane determination enable data variance to be captured and separated accordingly, providing better classification performance.

For quantitative assessment, best model accuracy from the several studies are presented in Table VII for studies using deep-learning approach to classify dense and non-dense breasts in mammography. Performance of the proposed improved model outperformed the others with best accuracy of 95.39%.

TABLE VII
SUMMARY ON PREVIOUS WORKS FOR BREAST DENSITY CLASSIFICATION USING MAMMOGRAM (BINARY CLASSIFIERS)

Author / Reference	Method	Best Model Accuracy (%)
Gandomkar et al. [7]	CNN	92.7
Mohamed et al. [11]	CNN	94.0
Lehman et al. [35]	CNN	88.0
Jian Deng et al. [36]	CNN + SE	92.17
Proposed method	**CNN + SVM**	**95.39**

CONCLUSION

In conclusion, the proposed improvement of the pre-trained model for extracting features of breast density regions in mammogram images with high variability between classes (i.e. CBIS-DDSM) needs a good classifier by introducing separating hyperplane through SVM to categorize the image better during classification. From the results, the overall system performance can be observed in GoogleNet+SVM, having the best accuracy of 95.39%, AUC of 0.9949 as well as highest sensitivity, specificity and precision scores. Furthermore, the training performance of GoogleNet also showed the best both qualitatively by observing its Grad-CAM heat map as well as quantitatively through its best validation vs training losses graph. Meanwhile, ResNet50+SVM also showed promising results when applied with the SVM classifier as it showed the most significant accuracy upgrade compared to using the original softmax classifier.

This study also demonstrated good performance of the system's generalizability when tested on other manufacturers' datasets. Although both datasets performed well within the same model, it is to note here that both datasets are from different manufacturer sources. Thus the dissimilarity may include settings that affect image quality, total radiation, and the manufacturer's 'post-processing' options. Lower performance observed might be affected by data variance in the training dataset (CBIS-DDSM), as well as class imbalance during training (lower dense image by two-third during training). As the CBIS-DDSM also includes lesions in non-dense images, this might cause the system to generalize it during training as a dense region and further lowers the training efficiency. This study is expected to be further expanded into testing the image from FFDM sources and datasets excluding abnormalities and tested for all four density classes. A publicly available dataset with segmented dense region ground truths is crucial for the further deep-learning-based algorithm to be developed. A suitable enhancement technique pertaining to the density information of the image might be explored in the future for the specific task of breast density classification as well as detecting masses appearance within the image.

ACKNOWLEDGMENT

This research work is financially supported by Ministry of Higher Education Grant Scheme (FRGS) (Ref: FRGS/1/2022/SKK06/UITM/02/3).The authors would like to express their gratitude to members of the Advanced Control System and Computing Research Group (ACSCRG), Advanced Rehabilitation Engineering in Diagnostic and Monitoring Research Group (AREDiM), Integrative Pharmacogenomics Institute (iPROMISE), and Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang for their assistance and guidance during the fieldwork. The authors are grateful to Universiti Teknologi MARA, Cawangan Pulau Pinang for their immense administrative and financial support. Special thanks to the Imaging Department, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Kepala Batas, Pulau Pinang for the professional consultation and expert guide.

REFERENCES

[1] S. M. Moss, C. Wale, R. Smith, A. Evans, H. Cuckle, and S. W. Duffy, "Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial.," *Lancet. Oncol.*, vol. 16, no. 9, pp. 1123–1132, Sep. 2015, doi: 10.1016/S1470-2045(15)00128-X.

[2] J. D. Picard, "[History of mammography].," *Bull. Acad. Natl. Med.*, vol. 182, no. 8, pp. 1613–1620, 1998.

[3] N. Wu *et al.*, "Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1184–1194, Apr. 2020, doi: 10.1109/TMI.2019.2945514.

[4] S. S. Nazari and P. Mukherjee, "An overview of mammographic density and its association with breast cancer," *Breast Cancer*, vol. 25, no. 3, pp. 259–267, 2018, doi: 10.1007/s12282-018-0857-5.

[5] K. M. Ray, E. R. Price, and B. N. Joe, "Breast Density Legislation: Mandatory Disclosure to Patients, Alternative Screening, Billing, Reimbursement," *Am. J. Roentgenol.*, vol. 204, no. 2, pp. 257–260, Jan. 2015, doi: 10.2214/AJR.14.13558.

[6] S. M. Advani *et al.*, "Association of Breast Density With Breast Cancer Risk Among Women Aged 65 Years or Older by Age Group and Body Mass Index," *JAMA Netw. Open*, vol. 4, no. 8, pp. e2122810–e2122810, 2021, doi: 10.1001/jamanetworkopen.2021.22810.

[7] Z. Gandomkar, M. Suleiman, D. Demchig, P. Brennan, and M. McEntee, *BI-RADS density categorization using deep neural networks*. 2019.

[8] A. Lee *et al.*, "BOADICEA: a comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors," *Genet. Med.*, vol. 21, no. 8, pp. 1708–1718, 2019, doi: 10.1038/s41436-018-0406-9.

[9] N. F. Boyd *et al.*, "Mammographic density and the risk and detection of breast cancer.," *N. Engl. J. Med.*, vol. 356, no. 3, pp. 227–236, Jan. 2007, doi: 10.1056/NEJMoa062790.

[10] Y. J. Suh, J. Jung, and B.-J. Cho, "Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning.," *J. Pers. Med.*, vol. 10, no. 4, Nov. 2020, doi: 10.3390/jpm10040211.

[11] 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.

[12] 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, 2019, doi: 10.7717/peerj.6201.

[13] N. F. Razali, I. S. Isa, S. N. Sulaiman, N. K. A. Karim, and M. K. Osman, "High-level Features in Deeper Deep Learning Layers for Breast Cancer Classification," in *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2021, pp. 170–175, doi: 10.1109/ICCSCE52189.2021.9530911.

[14] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020, doi: 10.1038/s41586-019-1799-6.

[15] Q. Zhang, Y. Li, G. Zhao, P. Man, Y. Lin, and M. Wang, "A Novel Algorithm for Breast Mass Classification in Digital Mammography Based on Feature Fusion," *J. Healthc. Eng.*, vol. 2020, p. 8860011, 2020, doi: 10.1155/2020/8860011.

[16] K. Chang *et al.*, "Multi-Institutional Assessment and Crowdsourcing Evaluation of Deep Learning for Automated Classification of Breast Density," *J. Am. Coll. Radiol.*, vol. 17, Jun. 2020, doi: 10.1016/j.jacr.2020.05.015.

[17] T. L. Nguyen *et al.*, "Novel mammogram-based measures improve breast cancer risk prediction beyond an established mammographic density measure.," *Int. J. cancer*, vol. 148, no. 9, pp. 2193–2202, May 2021, doi: 10.1002/ijc.33396.

[18] M.-L. Huang and T.-Y. Lin, "Considering breast density for the classification of benign and malignant mammograms," *Biomed. Signal Process. Control*, vol. 67, p. 102564, 2021, doi: https://doi.org/10.1016/j.bspc.2021.102564.

[19] W. B. Sampaio, E. M. Diniz, A. C. Silva, A. C. de Paiva, and M. Gattass, "Detection of masses in mammogram images using CNN, geostatistic functions and SVM.," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 653–664, Aug. 2011, doi: 10.1016/j.compbiomed.2011.05.017.

[20] I. Kumar, J. Virmani, H. Bhadauria, M. K. Panda, and D. Kriti, "Classification of breast density patterns using PNN, NFC, and SVM classifiers," in *Soft Computing Based Medical Image Analysis*, 2018, pp. 223–243.

[21] D. Schmidt *et al.*, "Cirrus: An Automated Mammography-Based Measure of Breast Cancer Risk Based on Textural Features," vol. 2, Dec. 2018, doi: 10.1093/jncics/pky057.

[22] M. Abdolell, K. Tsuruda, G. Schaller, and J. Caines, "Statistical Evaluation of a Fully Automated Mammographic Breast Density Algorithm," *Comput. Math. Methods Med.*, vol. 2013, p. 651091, 2013, doi: 10.1155/2013/651091.

[23] N. Saffari *et al.*, "Fully Automated Breast Density Segmentation and Classification Using Deep Learning.," *Diagnostics (Basel, Switzerland)*, vol. 10, no. 11, Nov. 2020, doi: 10.3390/diagnostics10110988.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[25] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An Automated In-Depth Feature Learning Algorithm for Breast Abnormality Prognosis and Robust Characterization from Mammography Images Using Deep Transfer Learning.," *Biology (Basel).*, vol. 10, no. 9, Sep. 2021, doi: 10.3390/biology10090859.

[26] M. Alkhaleefah and C.-C. Wu, "A Hybrid CNN and RBF-Based SVM Approach for Breast Cancer Classification in Mammograms," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 894–899, doi: 10.1109/SMC.2018.00159.

[27] S. A. Hassan, M. S. Sayed, M. I. Abdalla, and M. A. Rashwan, "Breast cancer masses classification using deep convolutional neural networks and transfer learning," *Multimed. Tools Appl.*, vol. 79, no. 41–42, pp. 30735–30768, Nov. 2020, doi: 10.1007/s11042-020-09518-w.

[28] J. O. Bandeira Diniz, P. H. Bandeira Diniz, T. L. Azevedo Valente, A. Corrêa Silva, A. C. de 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. 156, pp. 191–207, 2018, doi: https://doi.org/10.1016/j.cmpb.2018.01.007.

[29] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, no. 1, p. 170177, 2017, doi: 10.1038/sdata.2017.177.

[30] T. Cogan and L. Tamil, "Deep Understanding of Breast Density Classification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1140–1143, doi: 10.1109/EMBC44109.2020.9176628.

[31] J. Suckling *et al.*, "The Mammographic Image Analysis Society digital mammogram database," 1994.

[32] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, 2012, pp. 1097–1105.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[35] C. Lehman *et al.*, "Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation," *Radiology*, vol. 290, p. 180694, Oct. 2018, doi: 10.1148/radiol.2018180694.

[36] J. Deng, Y. Ma, L. Deng-ao, J. Zhao, Y. Liu, and H. Zhang, "Classification of breast density categories based on SE-Attention neural networks," *Comput. Methods Programs Biomed.*, vol. 193, p. 105489, Apr. 2020, doi: 10.1016/j.cmpb.2020.105489.

Noor Fadzilah Razali is currently pursuing her doctorate degree from Universiti Teknologi MARA (UiTM), Malaysia and working as a lecturer at the same University. She received her Masters Degree in Biomedical Engineering in 2014 and Bachelor Degree of Biomedical Engineering in 2010 from University of Malaya (UM), Malaysia. Her main research interests includes biomedical engineering, signal and image processing, and machine and deep learning application.

Iza Sazanita Isa received her bachelor's in electrical engineering from the Universiti Teknologi MARA, Malaysia in 2004 and the M.Sc. degree from the Universiti Sains Malaysia, Malaysia in 2008. Since 2009, she joined Universiti Teknologi MARA, Penang Campus, Malaysia as young lecturer and has been promoted as senior lecturer with the Faculty of Electrical Engineering, Universiti Teknologi MARA, in 2013. She pursues her PhD in Electrical Engineering under the SLAB/SLAI scholarship and graduated in 2018. Currently, she is attached to department of Control System Engineering at the faculty. She is members of AREDiM research group and the head of research group RIDyLT and actively involved in teaching and learning research.

Siti Noraini Sulaiman received her PhD degree in image processing from School of Electrical and Electronics Engineerin g, Universiti Sains Malaysia in 2012. She is currently an Associate Professor at School of Electrical Engineering, Universiti Teknologi MARA, Malaysia. She is now attached as the chair of Advanced Rehabilitation Engineering in Diagnostic and Monitoring (AREDiM) and the deputy chair of Advanced Control System and Computing Research Group (ACSCRG), Faculty of Electrical Engineering, Universiti Teknologi MARA, Penang Branch Campus in Permatang Pauh, Penang, Malaysia. Her main research interest is biomedical engineering focusing on image filtering, image clustering and algorithm for image processing.

Noor Khairiah A. Karim received the Bachelor Degree in Medicine, Bachelor Degree in Surgery (MBBS) and Master of Radiology (MRad) from the University of Malaya, Malaysia. She then obtained her Fellowship in Cardiac Imaging from the National Heart Center Singapore. She is currently a Senior Medical Lecturer of the Regenerative Medicine Cluster, and a Consultant Radiologist at the Advanced Medical and Dental Institute, Universiti Sains Malaysia. Her current research areas include medical image processing and analysis with special interest in cardiac, breast and brain imaging.

Muhammad Khusairi Osman obtained his B.Eng degree in Electrical and Electronic Engineering in 2000 and MSc in Electrical and Electronic Engineering in 2004 from Universiti Sains Malaysia. In 2014, he obtained his Ph.D. in medical electronic engineering from Universiti Malaysia Perlis (UniMAP), Malaysia. He is currently a senior lecturer at Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia. His research interest is in image processing, pattern recognition and artificial intelligence