A Hybrid CNN and RBF-Based SVM Approach for Breast Cancer Classification in Mammograms

Mohammad Alkhaleefah
Department of Electrical Engineering
National Taipei University of Technology
Taipei, Taiwan
t104999003@ntut.edu.tw

Chao-Cheng Wu
Department of Electrical Engineering
National Taipei University of Technology
Taipei, Taiwan
ccwu@ntut.edu.tw

***Abstract*—One of the most powerful ideas in deep learning is the transfer learning technique. Transfer learning can be utilized to take a knowledge from what a deep neural network has learned from a particular task and apply that knowledge to a different task. Transfer learning is very useful when the size of the training samples of interest is small to train a neural network from scratch. This research focuses on the concept of transfer learning where the Convolutional Neural Network (CNN) power can be utilized as a features extractor to help with classifying benign from malignant breast cancer images. In addition, Support Vector Machine (SVM) classifier based on Radial Basis Function (RBF) was adapted for its flexibility in fitting the data dimension space adequately by tuning the kernel width. The hybridization between CNN and RBF-Based SVM showed robust results for both the dataset and the application task of this research. The contribution of this paper can be summarized in three major parts. First, a CNN was implemented from scratch on a large number of available spine images to classify images of two different spine views (sagittal and axial) in order to transfer the learning process to the CNN of breast images. Second, retrain the spine CNN on the images of breast cancer to classify between benign and malignant cases by fine-tuning. Finally, the features were extracted from the retrained CNN and fed to RBF-Based SVM to classify benign from malignant breast mammogram images.**

***Keywords*—convolutional neural network, transfer learning, feature extraction, RBF-based support vector machine, breast cancer, classification, computer-aided diagnosis**

I. Introduction

Breast cancer is the most common cancer in women globally. In 2018, an estimated 266,120 new cases of malignant (invasive breast cancer) are expected to be diagnosed in women in the U.S.A, along with 63,960 new cases of benign (non-invasive breast cancer). Breast cancer can be defined as the abnormal growing of breast cells. Swelling, skin irritation, and a lump in the breast can be a symptom of breast cancer. However, many breast cancers have no obvious symptoms at all. Mammograms are probably the most important tool doctors use to diagnose and evaluate breast cancer [1].

Unfortunately, the false-positive rate (when a mammogram shows there is a cancer when in fact there is not) is high in mammography due to the structure of dense breasts. Dense breasts are harder to analyze through mammography than non-dense ones. In such cases, patients might need to perform an unneeded biopsy which can be costly, time-consuming, and uncomfortable [2].

Recently, Computer-Aided Diagnosis (CAD) systems have showed a high record of detection and classification accuracy of breast cancer in mammogram images, and they surpass the human performance in some cases, which can reduce the mortality rate among men and women with breast cancer.

Extracting and selecting relevant features for the traditional Machine Learning (ML) classifiers such as SVM is a very important stage. Nevertheless, if this stage is not being done properly, the overall accuracy and performance might be affected. Unfortunately, such classifiers need to be fed by hand-crafted features in order to perform a classification task.

Recent studies have showed that CNNs are capable of producing high accurate results on tasks such as image classification, object recognition, clustering, and segmentation in different fields of image processing. However, due to the large number of data needed to train a CNN from scratch, the transferability of different layers in pre-trained deep CNNs, such as AlexNet, GoogLeNet, OverFeat, and ResNet and reuse them for a new task, offered a promising opportunity to overcome the lack of training samples issue and showed great performance results [3].

A previous research [4] has showed how the transfer learning can be used to classify pre-detected breast masses from mammograms. It should be noticed that the model was proposed in [4] is an end-to-end deep learning model, which means the feature extraction and classification are all done by a CNN. Another study in [5] showed how the off-the-shelf Polynomial-Based SVM classifier was trained based on the features extracted from the pre-trained AlexNet CNN to distinguish between benign and malignant breast lesions.

The training samples used in these research papers [4], [5] are quite large compared to the training samples used in this

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00159

paper. In addition, the features extraction process was executed on the region of interest within the mammogram image. However, in this research, the features were extracted from the whole breast image and then the RBF-Based SVM classifier was fed with these features, not only to classify benign from malignant masses, but also to classify calcifications cases in breast mammograms.

Another work [6] used the features from a pre-trained OverFeat CNN that was trained for object detection in natural images, for nodule detection in computed tomography scans. In fact, the debate around how accurate the CNNs classification will be, when the original domain in which the CNN was trained varies from the domain of interest is still a controversy. This research tries to show that the domain similarity is important in the CNN transfer learning process. Therefore, a CNN was trained on the same domain as the domain of interest in this research.

This paper is organized as follows. Section II describes the methodology and materials of this research. Section III demonstrates the experiment results and performance evaluation methods. Section IV summarizes the conclusion and discussion.

II. METHODOLOGY AND MATERIALS

The idea behind neural nets transfer learning is to map the weights and biases learned from layers of one deep neural network to layers of a new neural network in order to make the learning process faster and hopefully producing more accurate results in the new neural network.

A. Convolutional Neural Network (CNN)

CNNs are made up of neurons that are arranged in 1, 2, or 3 dimensions way (width, height, depth) that transform the input values through a series of hidden layers to output class scores of neuron activations [7]. Unlike regular Neural Networks (NNs), the neurons in a layer will only be connected to a local area of the previous layer. The common layers in CNNs are CONV, FC, RELU, and POOL, in addition to the input and output layer [8]. The CNN layers architecture implemented in this research is briefly explained below.

- Input Layer: The input is an image of width 227, height 227, and channel size 1. The images were reduced to this size in order to compare the results of this research with other works later on.
- Convolutional Layer or CONV layer: In this layer, the filter numbers, the filter sizes, and padding are specified to compute the output of neurons that are connected to local regions in the input image [9]. In this CNN architecture, there are three CONV layers. The first CONV layer has 3 by 3 filter size and the number of filters is 16. The number of padding is 1. Fig. 1 illustrates the structure of the CONV layer and its connection with the input layer. Moreover, the number of filters or neurons is increasing through the deeper layers from 16 to 64 to compensate the increase in the abstractions need to be observed.

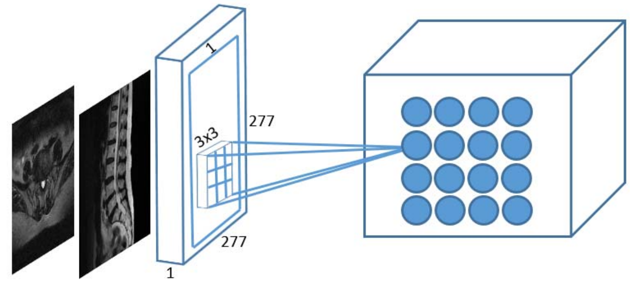

Fig. 1. The connectivity of the input layer (left) with the CONV layer (right), the input image size is 227x227x1, with 16 filters of size 3x3, and padding=1.

- Batch Normalization (BN) Layer: It is used between CONV layer and the non-linear Rectified Linear Unit function (ReLU) layer to speed up the network training. It is useful in reducing sensitivity to variations within the data [10].
- Rectified Linear Unit (ReLU) Layer: It is a nonlinear activation function which comes after the batch normalization layer to perform a threshold operation to each element of the input, where any value less than 0 is set to 0, and x for any value equal or greater than 0. This is equivalent to (1) [11].

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

- Pooling Layer: It is one way of performing a down-sampling operation. A down-sampling operation can reduce the spatial size of the features and remove the redundant spatial information that might be resulted from CONV layer [12].
- Fully Connected (FC) Layer: Each neuron in this FC layer is connected to the all active neurons in the previous layers. Therefore, FC layer combines all the features learned by the previous layers [12]. The number of the FC layers output must match the number of classes in the targeted data. Since there are two classes of spine images (sagittal and axial) views need to be classified, the output size of the FC layer parameter was set to 2.
- Softmax Layer: It normalizes the output of the FC layer, and outputs positive numbers that sum to one which can then be used to produce classification probabilities over the 2 class labels by the classification layer next. The output of the Softmax layer can be produced by Softmax activation function which is represented in (2). where y is an input vector, e^{y_i} is the output of the ith neuron, and e^{y_k} is different neurons in the softmax group [13].

$$\sigma(y)_i = \frac{e^{y_i}}{\sum_{k=1}^{K} e^{y_k}} \tag{2}$$

- Classification Layer: It is the final layer in the CNN architecture. It uses the probabilities returned by the Softmax activation function for each input to assign the input to one of the mutually exclusive classes and compute the loss. It is worth mentioning that the loss is computed by the cross entropy function for k mutually exclusive classes as shown in formula (3).

$$E(\Theta) = -\sum_{i=1}^{n}\sum_{j=1}^{k} t_{ji} \ln y_j(x_i, \Theta) \tag{3}$$

Where Θ is the parameter vector, tij is the indicator that the ith sample belongs to the jth class, and $y_j(x_i, \Theta)$ is the output for sample i. The output $y_j(x_i, \Theta)$ can be interpreted as the probability that the network associates ith input with class j, that is, $P(t_j = 1 \mid x_i)$ [14]. However, in this research the classification process is done by the RBF-Based SVM classifier as it will be described in the next section. However, the number of this CNN layers is an arbitrary number. It turns out that it is suitable for this research case.

After the brief description of the CNN, here is how it performs. The CNN neurons n_i compute a dot product of their weights w_i with the activation input x_i followed by a bias b to determine when the neuron weight needs to be active at, and all are wrapped by a ReLU function as shown in (4) [15].

$$n_i = f(\sum_{i=1} w_i x_i + b) \tag{4}$$

The goal of this paper was to implement an end-to-end CNN to classify benign from malignant cancer in mammograms, but unfortunately the training samples available are too small for this particular implementation. Therefore, since there are a large number of two different classes of spine images (axial and sagittal) views available, the CNN was trained on these images from scratch to classify axial from sagittal views. The reason behind this is to use the initiated CNN parameters and transfer them as a pre-training to the original task which is the breast cancer classification.

The end-to-end CNN has 15 layers including the input and output layer. The network took about 25 minutes to train on the spine images utilizing (NVIDIA Quadro P600 2G) GPU, whereas it took only 11 seconds to re-train the network on the breast images.

The spine CNN layers that have just been created have learned useful image features that can help to perform an image classification process. Next, these learned spine CNN layers will be transferred to another CNN for a breast cancer classification task. Since the breast cancer training images samples are small, the last 3 layers of the spine CNN (FC, Softmax, Classification) were fine-tuned, and the rest of the layers was copied to the new CNN of the breast cancer images.

Fig.2 illustrates the development of the features as they move from a layer to another and shows the part which was copied and the part which was fine-tuned. The last 3 layers of the spine CNN have learned the spine image views classifications, these layers need to be replaced by new layers that will learn to classify breast cancer images.

After fine-tuning the last 3 layers of the spine CNN, the network was re-trained on the new breast cancer dataset. Next, all the high level features in the FC layer of the re-trained CNN will be extracted and fed to the RBF-Based SVM classifier to classify benign from malignant breast cancer cases as shown in Fig.3.

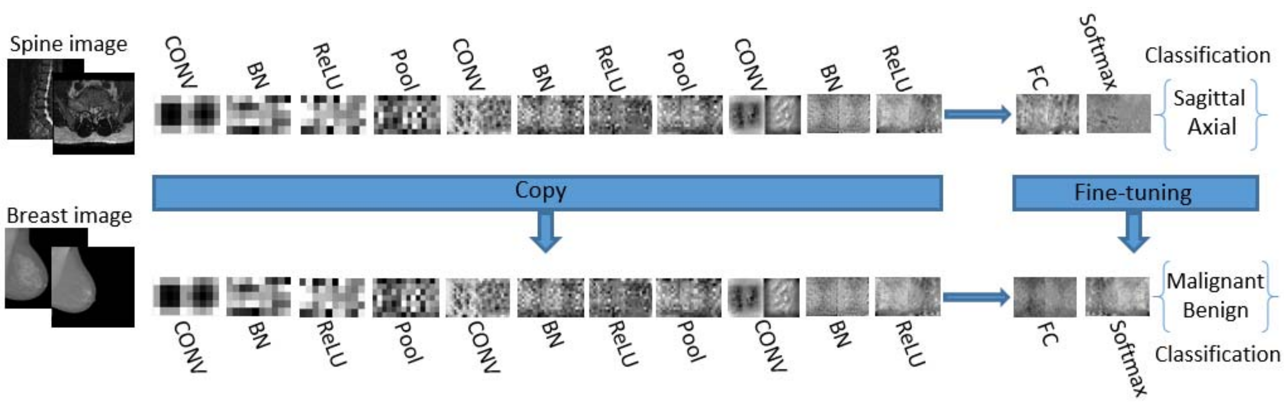

Fig. 2. The transformation between the Spine and Breast CNN, and the development of the layers within each CNN.

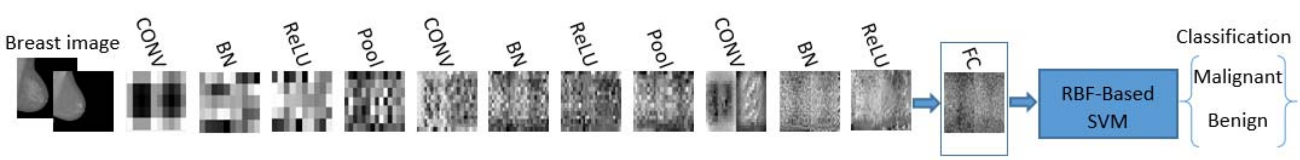

Fig. 3. Features extraction from FC layer to be fed into RBF-Based SVM classifier.

Table I shows the CNN training options. Where the Training Method corresponds to the learning algorithm technique for updating the CNN weights and biases, the Initial Learn Rate is used for training, Max Epochs are the maximum number of epochs that will be used for training, and the Mini Batch Size is used for each training iteration [16].

TABLE I
CNN TRAINING OPTIONS

CNN	Training Method	Initial Learn Rate	*Max Epochs*	*Mini Batch Size*
Spine	sgdm[a]	0.0001	30	32
Breast	sgdm	0.0001	80	32

[a] sgdm:Stochastic Gradient Descent with Momentum.

B. RBF-Based SVM Classifier

Support Vector Machine (SVM) has been used for many years in the field of machine learning for its classification power in both linearly and non-linearly separable data, especially for those problems that have small amount of dataset to train on, which is the case in this research. Due to the fact that most of the world problems is non-linear, providing more robust results, and compensating the weakness of the CNN in terms of the huge amount of training samples required to produce a good result, the RBF-Based SVM was used to classify benign from malignant cases.

RBF-Based SVM is a non-linear kernel function that can map the data from the input space to a higher dimensional feature space where data becomes linearly separable. The RBF-Based SVM function is shown in (5). Where x and y, represent feature vectors in an input space. $||x - y||^2$ is the squared Euclidean distance between x and y vectors, σ^2 is a kernel parameter [17].

$$K(x, y) = exp\left(- \frac{||x - y||^2}{2\sigma^2} \right) \tag{5}$$

The width parameter of the RBF kernel (σ) provides a flexibility in fitting the data samples and decide the decision boundary to distinguish between the benign and malignant cases. It is worthwhile to mention that the σ parameter with the best performance was chosen during the leave-one-out validation process on the training samples. Eventually, the features in the FC layer of the re-trained CNN were extracted by an activation function and fed to the RBF-Based SVM classifier to classify between benign and malignant cases.

C. Image Data

Two different dataset A and B were obtained from two different sources. The dataset A which contains 5120 MRI spine images of both views axial and sagittal was obtained from Taipei Medical University Hospital. The dataset B was acquired from (Centro Hospitalar de S. Jo ao [CHSJ], Breast Centre, Porto, Portugal). The dataset B has a total of 410 images, but only few biopsy proven cases were considered.

Therefore, 27 images of benign and malignant were used in this manuscript. The dataset A was used to train the first CNN from scratch in order to classify sagittal from axial view. The dataset B was utilized to fine-tune and retrain the first CNN to classify benign from malignant breast cancer. Table II represents the division of the dataset used in this manuscript. It is worth mentioning that only the biopsy proven mammogram breast images with all types of breast density were considered to provide a more robust classification.

TABLE II
THE DISTRIBUTION OF THE DATASET

Datasets	*Types*	*Train*	*Test*	*Class Total*	*Overall Total*
Dataset A (Spine)	Sagittal	2682	40	2722	5120
	Axial	2329	69	2398	
Dataset B (Breast)	Benign	8	7	15	27
	Malignant	6	6	12	

III. EXPERIMENTAL RESULTS

Leave-one-out validation was used to compute the misclassification rate and evaluate the hybrid approach before the testing process. The training data was partitioned into k=14 folds, training set and testing set. The number of folds is equal to the number of observations. Leave-one-out is a special case of k-fold cross-validation.

K-1 subsamples are used as training data, and 1 is used as the validation data for testing the hybrid approach. Finally, the results are averaged to produce the misclassification estimation. The misclassification rate is 0.21.

The most common evaluation methods in the field of medical imaging process were used for evaluating the performance of the proposed hybrid approach on the classification of benign from malignant cases. These evaluation methods are confusion matrix, sensitivity, specificity, and classification accuracy [18]. Each method is briefly explained below.

- Confusion Matrix: It represents information about the actual classes and the predicted classes of the hybrid approach classification. The actual class and the predicted class can be positive or negative. Table III represents the confusion matrix of the proposed hybrid approach. TN is the True Negative cases, when the actual case is benign and the predicted case is also benign. FN is the False Negative cases, when the actual case is malignant and the predicted case is benign. FP is the False Positive cases, when the actual case is benign and the predicted case is malignant. TP is the True Positive cases, when there is actual malignant case and the predicted case is malignant. It is worth to be noted that Negative (No) represents a benign cancer. Positive (Yes) represents a malignant cancer.
- Sensitivity: It is also called the TP rate. It is computing the number of true positive predictions over the number of actual positive cases. The calculation of sensitivity is formulated as shown in (6).

TABLE III
THE CONFUSION MATRIX OF THE HYBRID APPROACH

Number of samples=13	*Predicted: No*	*Predicted: Yes*
Actual: No	TN=6	FP=1
Actual: Yes	FN=0	TP=6

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

- Specificity: It is also called the TN rate. It computes the proportion of actual negative cases that are predicted as negative cases. Formula (7) presents the specificity measurement.

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

- Accuracy: It represents the number of correct predicted cases over the all cases. It can be formulated as in (8).

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \quad (8)$$

The accuracy of the proposed hybrid approach is compared with the best models in [4], [5] where the transfer learning was explored had used the same base architecture as the pretrained networks of AlexNet and GoogLeNet but fine-tuning on mammography images. However, the pre-trained network of this research has been trained on 5120 images which is a fairly small dataset compared with AlexNet and GoogLeNet which were trained on more than a million of natural images and can classify images into 1000 object categories. In addition, in this research, AlexNet and GoogLeNet were fine-tuned and tested on the same dataset used in this research in order to make a fair comparison between those approaches and the hybrid approach. Table IV shows the performance evaluation comparison between the different approaches.

TABLE IV
EVALUATION COMPARISON

Method	*Sensitivity*	*Specificity*	*Accuracy*
AlexNet-Polynomial-Based-SVM	0.83	0.86	0.85
GoogLeNet-CNN-Softmax	0.83	0.71	0.77
Hybrid-Approach	1	0.86	0.92

IV. CONCLUSION AND DISCUSSION

This research introduces a possible solution to classify benign from malignant breast cancer in mammograms by a hybrid approach which combines a CNN and RBF-based SVM. This approach applies the concept of transfer learning which can be very useful specially when there is a lack of medical images to train a CNN from scratch to perform a particular task. In addition, the hand-crafted features phase that is needed by the traditional machine learning classifiers such as SVM was avoided. Instead, the features were automatically generated by the CNN and fed to the RBF-based SVM classifier to classify between benign and malignant breast cancer images.

The RBF-based SVM classifier flexibility in terms of fitting the data samples space by controlling the width parameter of the RBF kernel (σ) and producing robust results are the main reasons behind selecting the RBF-based SVM. It is believed that the transfer learning between two CNNs that are similar in the domain structure is more effective than those with different domains. Therefore, the hybrid approach outperforms those models that relied on the domain-different pretrained CNNs such as AlexNet and GoogleNet.

Implementing this hybrid approach went through three phases, phase 1, since there are not enough mammogram breast images to train the CNN from scratch, instead the CNN was trained on spine MRI images slices to classify between two views of the spine (sagittal and axial) views. In phase 2, this spine CNN was fine-tuned and retrained on a few of the benign and malignant breast images. During the first phase, all the parameters within the spine CNN were initiated with different values, which means the network has learned to make an image classification task. In phase 3, all the learned features in the fully connected layer of the re-trained CNN will be extracted and fed to the RBF-Based SVM classifier to classify benign from malignant cases.

The evaluation methods used to evaluate the performance of the hybrid approach are confusion matrix, sensitivity, specificity, and classification accuracy. The hybrid approach outperforms the common approaches used in other research papers such as AlexNet and GoogleNet pre-trained networks.

ACKNOWLEDGMENT

This research was supported by Taipei Medical University Hospital and University Hospital (Centro Hospitalar de S. Jo ao [CHSJ], Breast Centre, Porto, Portugal).

REFERENCES

[1] T. A. C. S. (2018). Cancer facts and figures 2018. Accessed 2018-03-21. [Online]. Available: https://goo.gl/wHFXhf

[2] K. Kerlikowske, W. Zhu, R. A. Hubbard, B. Geller, K. Dittus, D. Braithwaite, K. J. Wernli, D. L. Miglioretti, E. S. OMeara, B. C. S. Consortium *et al.*, "Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy," *JAMA internal medicine*, vol. 173, no. 9, pp. 807–816, 2013.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[4] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," *arXiv preprint arXiv:1612.00542*, 2016.

[5] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.

[6] B. van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 286–289.

[7] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*. O'Reilly, 2017.

[8] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 248–251.

[9] S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize alzheimer's disease using fmri data," in *Future Technologies Conference (FTC)*. IEEE, 2016, pp. 816–820.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[11] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014, pp. 844–848.

[12] Z. Yu, E.-L. Tan, D. Ni, J. Qin, S. Chen, S. Li, B. Lei, and T. Wang, "A deep convolutional neural network based framework for automatic fetal facial standard plane recognition," *IEEE journal of biomedical and health informatics*, 2017.

[13] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Computer Methods and Programs in Biomedicine*, 2018.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[15] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "Xnor-net: Imagenet classification using binary convolutional neural networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 525–542.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] T. Mu and A. K. Nandi, "Breast cancer detection from fna using svm with different parameter tuning systems and som–rbf classifier," *Journal of the Franklin Institute*, vol. 344, no. 3-4, pp. 285–311, 2007.

[18] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert systems with applications*, vol. 36, no. 2, pp. 3240–3247, 2009.