Breast Cancer Detection Based on Deep Learning Technique

Nur Syahmi Ismail
Department of Electrical & Electronics Engineering
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750
Tronoh, Perak, Malaysia
syahmi_17007393@utp.edu.my

Cheab Sovuthy
Department of Electrical & Electronics Engineering
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750
Tronoh, Perak, Malaysia
sovuthy.cheab@utp.edu.my

Abstract- **Breast cancer is the most common cancer among Malaysian women and roughly one in 19 women at risk of breast cancer in Malaysia. The number of breast cancer cases is steadily growing especially with increasing number of ageing population. The screening practice using mammography needs to be better and potentially efficient. There is always room for advancement when it comes to medical imaging. Early detection of cancer can reduce the risk of deaths for cancer patients. The objective of this paper is to compare the breast cancer detection with two model networks of deep learning technique. The overall process involves image preprocessing, classification and performance evaluation. In this paper, we evaluate the performance of deep learning model network which are VGG16 and ResNet50 to classify between normal tumor and abnormal tumor using IRMA dataset. The result show that VGG16 produces the better result with 94% compared to ResNet50 with 91.7% in term of accuracy.**

Keywords—breast cancer, mammogram, deep learning, transfer learning

I. INTRODUCTION

Breast cancer [1] starts when abnormal cells in the breast begin to grow uncontrollably. Two common types of breast cancer are ductal cancer and lobular cancer. Ductal cancer starts from the duct to the nipple and lobular cancer starts in gland that produce breast milk. Risk factor for most cases of breast cancer is due to a combination of factors which are genetic and environmental [2]. The risk factors that we cannot change or control is genetic factors such as family history, early menstruation, late menopause and dense breast tissue [3]. The example of environmental and lifestyle risk factors that we have control of are being overweight or obese, lack of physical activity and drinking alcohol. Based on Malaysian National Cancer Registry Report (MNCR) 2007-2011 published by National Cancer Institute [4], breast cancer is the number one cancer for women in Malaysia with 18,343 patients (17.7%) followed by colorectal cancer with 13,693 patients (13.2%). The occurrence of breast cancer was the highest among Chinese followed by Indian and Malay. The challenge in handling breast cancer is to be able to provide comprehensive service in diagnosis and treatment. One of the commonly used imaging modalities for early breast cancer detection is mammography in which an abnormality can be categorized as either normal, non-cancerous (benign) or cancerous (malignant) [5]. The ability to stop breast cancer by detecting it early can help to increase the percentage of patients that can survive. During a mammogram screening, X-ray images are captured from two angles of each breast. Screening mammograms are evaluated and inspected by human readers, the experienced radiologist [6]. Double reading was found to improve the result of evaluation and it has been applied in many countries [7]. Multiple reading up to more than 10 readers can further improve the diagnostic performance. There is plenty of room for improvement in mammogram evaluation beyond the double reading.

Numerous studies have been conducted to evaluate tumor detections to mammography images. Several researchers have used traditional texture analysis [8], extreme learning machine (ELM) [9] and random forest classifier [10] to classify the mammogram image from the selected dataset to either normal and abnormal tumor or normal, benign and malignant.

Other researchers have used IRMA dataset and classify the mammogram using histogram oriented gradient (HOG) [11],[12] and local configuration pattern (LCP) [13].

This paper presents a study between two model networks which are VGG16 and ResNet50 for classification of mammogram from IRMA dataset.

II. METHODOLOGY

Fig. 1 shows the block diagram of overall breast cancer detection system. Raw images collected from Image Retrieval in Medical Application (IRMA) dataset are pre-processed using image resize and image conversion to fit the network system. Next, classification using two model networks, VGG16 and ResNet50 are implemented. All the images are classified based on normal tumor or abnormal tumor. The performance of classification is measured using performance evaluation metrics such as precision, recall and accuracy.

978-1-7281-3494-9/19/$31.00 ©2019 IEEE

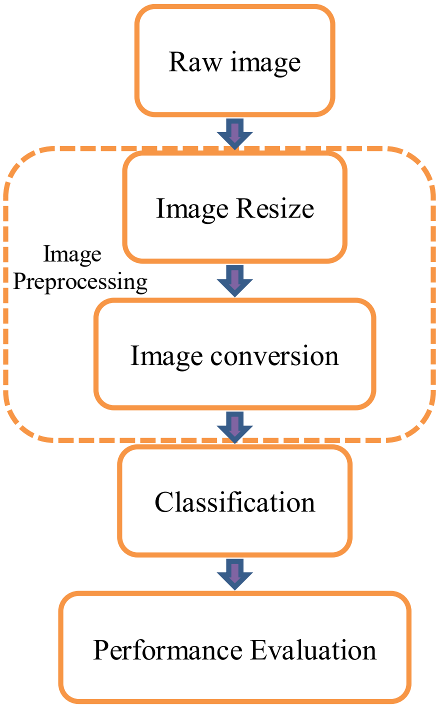

Fig. 1. Block diagram of breast cancer detection system

A. Dataset

This research will be using mammography images from public datasets of the IRMA database from RWTH Aachen, Germany. IRMA is a standard database that provides mammogram patches. The available mammography database developed from the four different datasets which are Mammographic Image Analysis Society Digital Mammogram Database (MIAS), Digital Database for Screening Mammography (DDSM), Lawrence Livermore National Laboratory (LLNL), and routine images from the Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen. The final size of mammogram images patches are 128 × 128 pixels where 931 images are normal and 584 are abnormal (benign and malignant). Sample patches of each class of the IRMA database are shown in Fig. 2. The rows of the Fig. 2 show samples of the normal and abnormal cancer cases.

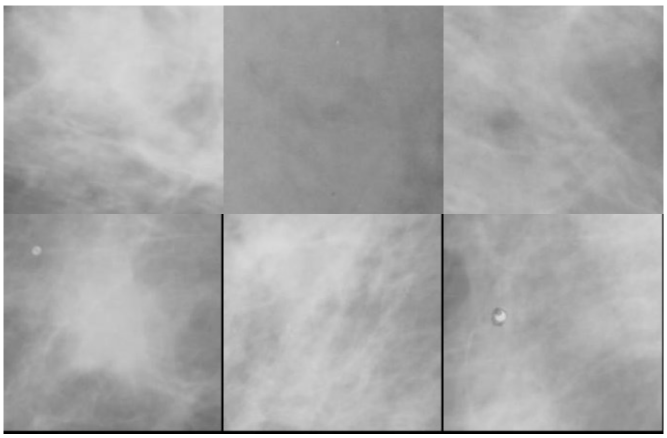

Fig. 2. Sample of mammogram patches

B. Image preprocessing

i) Image resize

A resizing operation is applied to fit the images into the input layer of the network. The raw data pixel size is too small. The network needs an image input size of 224-by-224. These images are resized to 224 x 224 for uniform dimension as input image of the classifier.

ii) Image conversion

Some of the mammogram images ae not 3-channel input image. Standardized the images to have 3-channel input image needed for the network to function. The grayscale images are converted into RGB images for the 3-channel input images.

The dataset is split into training and validation data. 30% of the images is set for the training data and 70% for the validation data. To avoid the results being biased, the split is randomized.

C. Classification

Convolutional Neural Networks (CNN) are a category of neural network that is so far has been most popularly used in application of image recognition and computer vision. It has been proven very effective in areas such as image recognition and classification. Just like other Neural Networks, CNN consists of an input layer, multiple hidden layers and an output layer. Between the convolution layers often lies pooling layers that perform subsampling on the data to reduce training overheads. Each convolution layer has many filters, the size of which is smaller than the input, that independently perform the convolutions across the image. These filters learn patterns across the entire image. As the input is passed through the network, the convolution layers perform convolutions on the image. In convolution layers, the neurons will only connect to a limited region of the previous layer. This reduced computation complexity, and enables CNN to make full use of the 2D structure of the input data. Therefore, compared to other deep learning architecture, CNN can often lead to better result in image or speed recognition [14]. The basic architecture of CNN as shown in Fig. 3.

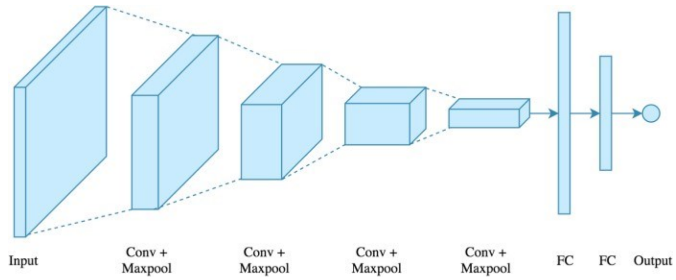

Fig. 3. Architecture of CNN

978-1-7281-3494-9/19/$31.00 ©2019 IEEE

Transfer learning is used in majority cases in deep learning because it able to train the deep neural network with comparatively little data. In computer vision, neural network detected different information from all layers. It helps to leverage the labeled data of task that train initially. Pre trained network can be used as a first step to learn a new task. Pre-trained CNN alongside fine-tuningand transfer learning lead to faster convergence and outperform training from scratch [15].

The dataset used in this project is smaller than the used reference dataset (ImageNet; training data with 1.2M) [15]. The weight of different layers is initialized for the proposed network using VGG16 and ResNet50 pre-trained models. The first layer of the network learned filters for capturing blob and edge features. These features are then processed by deeper network layers, which combine the early features to form higher level image features. Then, the last layer fine-tuning on cancer images dataset and their labels.

1) Models of CNN:

a) Visual geometry group (VGG): VGG-16 is a CNN model that is trained on the ImageNet database that has over a million images. The network is 16 layers deep and able to classify images into 1000 object categories. VGG16 has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224 [16]. The architecture of VGG16 as shown in Fig.4.

b) ResNet: Microsoft's ResNet known for its depth (152 layers) and the introduction of residual blocks. The residual address the problem of training a really deep architecture by introducing identity skip connection so the layers can copy their input to the next layer [16]. The architecture of ResNet50 as shown in Fig.5.

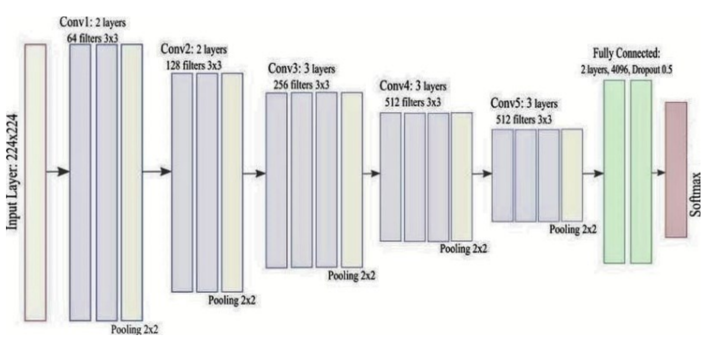

Fig. 4. Architecture of VGG16

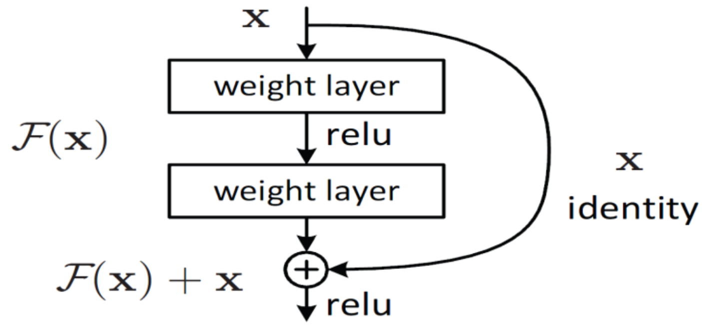

Fig. 5. Architecture of ResNet50

D. Classification performance

Classification performance is evaluated based on the following.

a) Precision: Precision measures the ability of the system to correctly identify the class of breast cancer in term of true negative (TN) and false positive (FP).

$$\text{Precision} = \frac{TN}{FP+TN} \tag{1}$$

b) Recall: Recall measure the performance of classification of system to identify the breast cancer class in term of true positive (TP) and false negative (FN).

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

c) Accuracy: It measures the accuracy of classification.

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

III. RESULT AND DISCUSSION

The result for the project is the cancer type's classification which are normal and abnormal tumor. Table 1 show the standard performance metrics calculated by the confusion matrices. Table 1 and Figure 6 signify a comparison of overall performance evaluation breast cancer detection from IRMA dataset classification using the VGG16 network and ResNet50 network, and previous research work done by Q. Zhang et. al [11] that classify the IRMA dataset by using Histogram Oriented Gradient (HOG).

Table 1. Comparison overall performance between different method of classification of IRMA dataset

	VGG16	**ResNet50**	**Q. Zhang et. al [11]**
Precision	89%	88%	82%
Recall	99%	94%	86%
Accuracy	94%	91.7%	83.2%

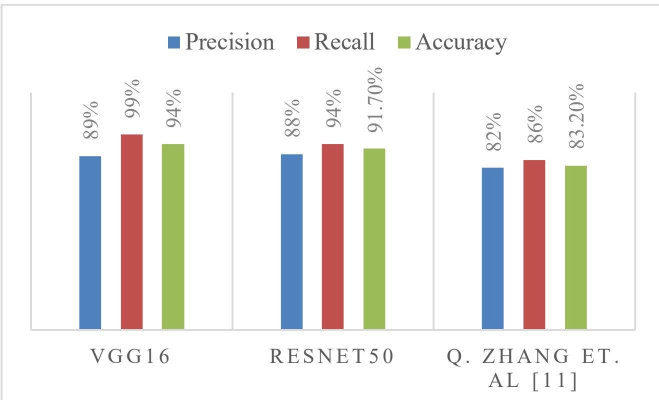

Fig. 6. Comparison performance evaluation between VGG16, ResNet50 and Q. Zhang et. al [11]

Table 1 and Fig. 6, illustrate the overall performance of VGG16 that indicates the better result compared to ResNet50 and Q. Zhang et. al [11]. The result of precision shows that VGG16 is 89% as compared to ResNet50 with 88% and Q. Zhang et. al [11] with 82%. The result of recall shows that VGG16 give the better result which is 99% as compared to ResNet50 with 94% and Q. Zhang et. al [11] with 86%. The result of accuracy shows that VGG16 gives 94% while ResNet50 is 91.70% and Q. Zhang et. al [11] with 83.20%.

The layers of the network are different as VGG16 has 16 layers while ResNet50 has 152 layers. The numbers of layers of the architecture affect the process time but with larger capacity that the network has and also it can cause overfitting.

IV. CONCLUSION

In this paper, deep learning technique using VGG16 and ResNet50 network have been implemented for normal and abnormal breast cancer detection. The classification methods were evaluated using three performance evaluations which are precision, recall, and accuracy rate. The best result of classification accuracy was VGG16 with 94%. For future work, the abnormal images can be classified to malignant and benign tumor. That help a lot in term of conducting the next procedure for the patients.

REFERENCES

[1] https://www.cancer.org/cancer/breast-cancer/about, downloaded 20 February 2019
[2] https://www.nationalbreastcancer.org/what-is-breast-cancer, downloaded 20 February 2019
[3] O. Golubnitschaja, M. Debald, K. Yeghiazaryan, W. Kuhn, M. Pešta, V. Costigliola, and G. Grech, "Breast cancer epidemic in the early twenty-first century: evaluation of risk factors, cumulative questionnaires and recommendations for preventive measures," Tumor Biology, vol. 37, no. 10, pp. 12941–12957, 2016.
[4] Malaysian National Cancer Registry Report
[5] N. Houssami, C. I. Lee, D. S. M. Buist, and D. Tao, "Artificial intelligence for breast cancer screening: Opportunity or hype?," *Breast*, vol. 36, pp. 31–33, 2017.
[6] M. S. Bae *et al.*, "Breast Cancer Detected with Screening US: Reasons for Nondetection at Mammography," *Radiology*, vol. 270, no. 2, pp. 369–377, 2013.
[7] D. Wang and A. Khosla, "Deep Learning for Identifying Metastatic Breast Cancer," pp. 1–6,2016.
[8] S. K. M. Hamouda, R. H. A. El-ezz, and M. E. Wahed, "iMedPub Journals Enhancement Accuracy of Breast Tumor Diagnosis in Digital Mammograms Keywords," pp. 1–8, 2017.
[9] W. Xie, Y. Li, and Y. Ma, "Neurocomputing Breast mass classi fi cation in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930–941, 2016.
[10] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated Mass Detection in Mammograms using Cascaded Deep Learning and Random Forests." pp. 1-8, 2015.
[11] Q. Zhang, I. U. Haq, A. Jadoon, A. Basit, and S. Butt, "Classification of mammograms for breast cancer detection based on curvelet transform and multi-layer perceptron.," vol. 28, no. 10, pp. 4311–4315, 2017.
[12] A. A. Shastri, D. Tamrakar, and K. Ahuja, "Density-Wise Two Stage Mammogram Classification using Texture Exploiting Descriptors ☆," no. 25, 2018.
[13] İ. I. Esener, S. Ergin, and T. Yüksel, "A New Ensemble of Features for Breast Cancer Diagnosis," *2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron.*, no. May, pp. 1168–1173, 2015.
[14] D. Bc-droid, R. Platania, K. Lee, and S. Park, "Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID)," pp. 536-543, 2017.
[15] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion : Understanding Transfer Learning for Medical Imaging," 2019.
[16] A. Soudani and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," *Expert Syst. Appl.*, vol. 118, pp. 400–410, 2019.