Convolutional Autoencoder Application for Breast Cancer Classification

Maryam Naderan
Department of Mathematical Methods of System Analysis
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
ma.naderan@gmail.com

Yuri Zaychenko
Department of Mathematical Methods of System Analysis
Igor Sikorsky Kyiv Polytechnic Institute
Kyiv, Ukraine
zaychenkoyuri@ukr.net

***Abstract*—There are many related works for breast cancer detection using convolutional neural networks (CNN), and most of them rely only on feature extraction after the convolutions. However, because of the huge number of parameters in the models, the time of computation will be increased. In the current work, a convolutional autoencoder was proposed to reduce the complexity of the model, the number of parameters and as a result, prevent overfitting the model. In this paper, there were trained several autoencoder models from scratch with different architectures and different hyperparameters. Moreover, 37% noise was applied on the input images and the model could reconstruct the original image with 84.72% accuracy. The sensitivity and precision of the proposed model were 86.87% and 80.23% respectively.**

Keywords*—*Breast Cancer Detection, Convolutional Autoencoder, Image Classification, DenseNet Network, Deep Learning Algorithm

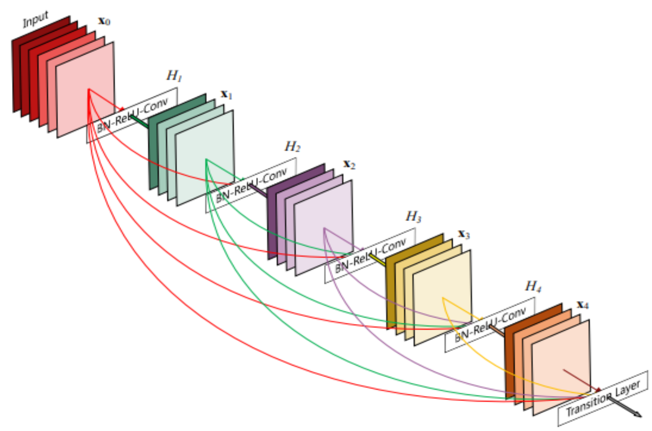

Fig. 1. A 5-layer dense block

I. Introduction

Breast cancer is a very common cancer among women between the ages of 35 and 55 [1]. Diagnosing breast cancer is frequently discussed as a classification problem within neural networks. Detecting and diagnosing breast cancer in early stages is critical in saving women's lives. Detecting this cancer in its early stages can help prevent the spread of cancer to other organs/tissues allowing doctors to help the patient before its too late. Early detection requires methods that are systematic and dependable, allowing healthcare professionals to accurately distinguish between benign and malignant tumors [2]. For these reasons, the exact detection and classification of breast tumors is extremely important for public health and to the lives of cancer patients [3].

There are four types of breast cancer: in situ, invasive ductal carcinoma, inflammatory breast cancer, and metastatic cancer [4]. Breast cancer detection is important in developing countries, where the number of patients is dramatically higher. Moreover, detecting breast cancer is a challenging and time-consuming task requiring doctors to manually label scans. Although, there are supervised and unsupervised machine learning algorithms that assist doctors. According to the World Health Organization (WHO) [5], a mammography scan is more efficient and cost-effective. Although it is more expensive than other medical images, the quality of the image is superior to other medical scans and therefore mammography scans, from the BreakHist dataset have been used in this work.

The main goal of this work is the development and investigation of deep learning algorithms to reduce the complexity of the model for breast cancer detection. A convolutional autoencoder was proposed to extremely decrease the computation time.

II. Review of previous works

There are a lot of studies that consider breast cancer detection using CNNs. However, most of them rely on the accuracy in their experiments, but accuracy in any cancer detection is not the only valid factor that should be considered [6]. In these tasks, the sensitivity of models should be considered to understand how many times the model misclassified cancer. Authors in [7] proposed a state-of-the-art convolutional neural network (DenseNet) for breast cancer detection using Breast Cancer Histology images (BACH) with an accuracy of 85.6%. The misclassification rate for cancer class was 14.4% on average. In their work, the sensitivity (on average for 4 classes)

978-1-7281-9083-9/20/$31.00 ©2020 IEEE

of ResNet 50 was compared with their proposed CNN at 76% and 79% respectively.

Compared with pre-existing CNN models (VGG-16, VGG-19, Xception, Resnet, Inception) with 80% accuracy in multi-class classification, authors in [8] proposed a model where the accuracy was 83.97% on average for two classes (Benign and Malignant). The proposed model was a combination of Inception and Resnet using the BreakHist data set, which contains 7909 mammography scans with four magnification zooms (40X, 100X, 200X and 400X).

In [9], it was stated that because of the architecture of DenseNet, in which all layers are fully connected to every previous layer, and with a short connection between those layers near the input and output, the model could be trained more efficiently and accurately. A brief structure of the DenseNet block is illustrated in fig.1. [9]

In [10] a DenseNet network was proposed with a training time of 11 hours. The Authors claim they first used weights from Imagenet and fine tuned the model to train DenseNet. All convolutional parts of the network were frozen but the fully connected layer was trainable. Authors in [11] used an atrous DenseNet that achieves multi-scale feature extraction by integrating the atrous convolutions to the dense block. The authors in [11] compared two datasets, BACH and CCG, in which the average class accuracy for the proposed model was 82.50% and 87.05% respectively for each dataset.

A new model of convolutional neural network was proposed in [12], where the authors used 400 images with 40x magnification for training data and 200 for validation data. In [12] three different ConvNet architectures were evaluated: 1) a 3-layer ConvNet architecture, 2) a 4-layer ConvNet architecture, and 3) a deeper 6-layer ConvNet architecture. The 3-layer ConvNet included one convolution, one pooling and one fully connected layer. The 4-layer had two convolutional and two pooling layers and the last layer was fully connected. The 6-layer ConvNet architecture comprises four convolutional and pooling layers with 16 units, a fully connected layer. According to the results in [12], deep architectures shows better result with 1.06% accuracy.

Authors in [13] proposed semi-supervised learning (SSL) using convolutional neural networks. The accuracy of the developed model was 82.43% and the area under the curve (AUC) observed in their study was 88.18%. There were 1874 pairs of mammogram images used during the experiments. Moreover, the authors developed three data weighing equations using exponential function, Gaussian function, and Laplacian function. Based on results [13], comparing two other weighting equation, the exponential function has shown better results with 82.43%, 81.00% and 72.26% for labeling accuracy, sensitivity and specificity respectively.

In [14] authors applied Principal component analysis (PCA) for Hybrid Fuzzy CNN Network. The idea of using PCA was to reduce the number of extracted features. In their work, the authors proposed a model where CNN VGG 16 was used for feature extraction and FNN NEFClass was used for image classification.

III. Dataset

The open source BreakHist dataset was used during the experiment. The dataset includes two classes benign and malignant. The dataset is also separated into four magnification zooms 40X, 100X, 200X and 400X. 5000 images were used for training and 350 images were used for testing. Figure 2 illustrates some input images that were used for training the model. Figures 2.a - 2.d belong to the benign category and Figures 2.e - 2.h belong to the malignant category.

IV. Architecture and Training of Convolutional Autoencoder

The aim of the autoencoder is to learn a compressed distributed representation for the given data typically for the purpose of dimensionality reduction. On the other hand, there is a principal component analysis (PCA) for the same task (reduction dimensionality). However, there are some advantages of using autoencoder like: 1) autoencoder can represent both linear and non-linear transformations in encoding but PCA can perform only linear transformations; 2) it could be more efficient in terms of model parameters to learn several layers with an autoencoder rather than one massive transformation with PCA; 3) it gives a representation as the output of each layer and having multiple representation of different dimensions is more practical.

One of the reasons a convolutional autoencoder was used during experiments is because it is very challenging to find

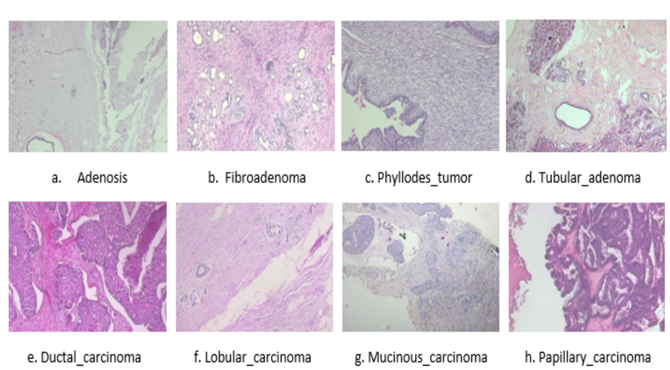

Fig. 2. Sample of input images

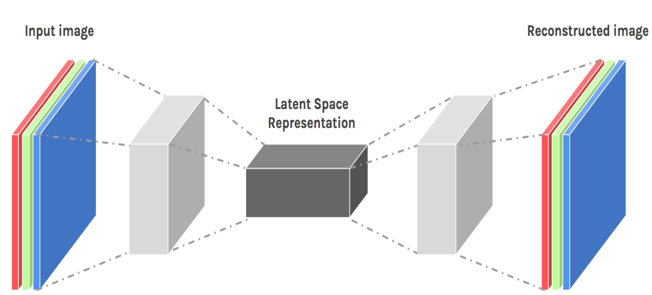

Fig. 3. The architecture of an autoencoder

significantly sized datasets with labels and autoencoder is an unsupervised model that does not require the dataset to be labeled. Another advantage of the autoencoder is that it makes the model smaller. Respectively, the model would have less parameters and as a result, the time of computation and training will drastically decrease. For example, in DeneseNet there are a total of 58,420,802 parameters and 7,037,504 of them are not trainable. However, in the proposed convolutional autoencoder there are 2,940,865 parameters and only 3,840 of them are non-trainable.

Figure 3 illustrates the architecture of an autoencoder. In the autoencoder there are layers between the input and output and the sizes of these layers are smaller than the input layer. For example, the input vector has a dimensionality of N which means that the output will also have a dimensionality of N. The input goes through a layer of size P, where the value of P is less than N. The autoencoder receives unlabeled input which is then encoded to reconstruct input. The important part of autoencoder is the Bottleneck approach for representation learning.

In the current work, several architectures of convolutional autoencoder were used during the experiment. The convolutional autoencoder was modified with 18 encoding layers and 14 decoding layers. There were eight convolutional and two maxpooling layers in encoder. In decoder there were six convolutional and two upsampling layers. Batch normalization was used between each convolutional layer. The proposed convolutional autoencoder was trained in a way, that the model would extract informative features (Codes) during the encoding process, and the decoder could then reconstruct the original input image of the encoder. The model could recreate the original image, even though some noises were applied to the scans. A comparison of the input images and reconstructed images, is shown in figure 4. After creating a successful autoencoder-model, the output of the encoder will be used with a fully connected layer to create a full model (Convolutional Autoencoder).

Figure 5 (a) illustrates a sample of testing data without noise added and 5 (b) shows the result of convolutional autoencoder. The accuracy of the recreated test data for convolutional autoencoder was 79.38%.

TABLE I. RESULT OF COMPARISON DIFFERENT METHODS FOR DETECTING BREAST CANCER

Methods	Factors				
	Precision	***Recall***	***F1_score***	***Accuracy***	***Training_Time***
Convolutional Autoencoder	90.40%	90%	90.20%	90%	12h
DenseNet FT	91%	90%	90.50%	95%	19h
DenseNet FS	70%	65%	62.66%	67.50%	24h

V. EXPERIMENTAL INVESTIGATIONS AND ANALYSIS

For training an autoencoder there are four parameters that it is needed to be set. The first one is code size. The code size represents the number of nodes in the middle layer and smaller size results in more compression. The second parameter is the number of layers and the autoencoder could be as deep as we want it to be. Another parameter is the loss function. The last parameter is the number of nodes per layer. The number of nodes per layer decreases with each subsequent layer of the encoder and increases back in the decoder. Also, the decoder is symmetric to the encoder in terms of layer structure.

The Adam optimizer with learning rate 0.001 was used for training DenseNet whereas, in convolutional autoencoder the RMSprop(lr=0.001) has shown better results.

All scans were pre-processed, before being used to train the model by resizing, normalizing and dimensionality reduction methods. In this paper, all experiments were developed using Jupyter Labs, Tensorflow 2 and Python 3. The programs were implemented on a virtual machine with an NVIDIA Tesla GPU and eight Intel CPUs.

Figure 6.a and 6.b illustrate how the loss for training and validation data was changed. Multiple tests were done using different numbers of epochs. It was found that 250 epochs provided better results compared to larger number of epochs like 500. The more epochs, the more chance the model will have overfitting. According to figure 6, it is better to use learning rate 0.001 for the current model. However, this parameter could be different for other models.

In previous work [15], a modified Inception V3 was proposed for breast cancer detection. In this work, a DenseNet121 was fine-tuned and proposed modified convolutional autoencoder. Table 1 shows three different deep learning algorithms that were used for the current task. According to table 1, the

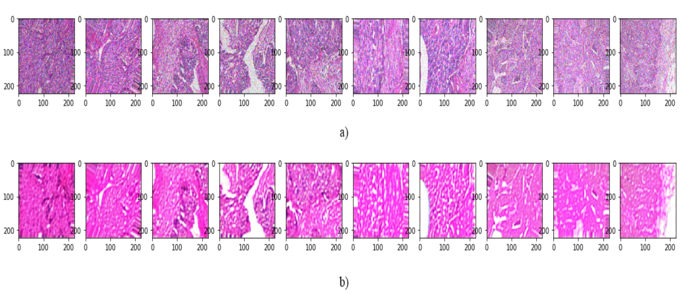

Fig. 4. On scans (a) noises were added, and scans (b) are reconstructed to original test data

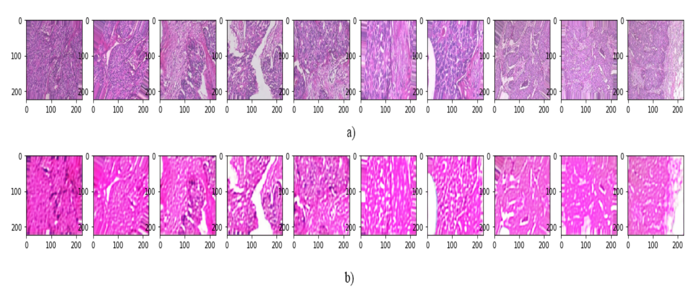

Fig. 5. (a) actual test data and (b) are reconstructed scans

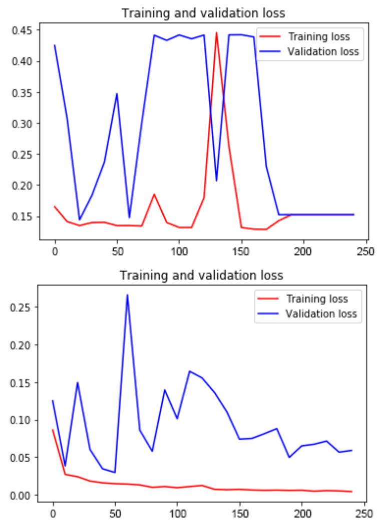

Fig. 6. Loss comparison for training and validation data with learning rate 0.01 and 0.001 on (a) and (b) respectively.

modified convolutional autoencoder has shown better results and compared to other methods, the training time was lower. There are plenty of works for breast cancer detection using deep learning, and most of them rely only on accuracy. Whereas in this paper, sensitivity (recall) and precision are considered in addition to accuracy. In the case of False Positive the outcome would be further investigation by medical professional. Whereas, in the case of False Negative the outcome would lead to a delay in proper diagnosis of the cancer. For this reason, sensitivity is an extremely important measurement in the effectiveness of the convolutional model. With sensitivity it's possible to assess whether the network predicts cancer as a cancer.

Fig. 7 illustrates a comparison between different methods of machine learning algorithms. According to fig.7, even though the accuracy and precision of DenseNet [15] was better than convolutional autoencoder by 0.88% and 1.93% respectively, there was a significant better result of sensitivity for the autoencoder model.

VI. Conclusion

In this paper, a convolutional autoencoder was developed, and the accuracy, sensitivity and precision of the proposed model were 84.72%, 86.87% and 80.23% respectively.

The convolutional autoencoder has fewer parameters compared to DenseNet, as a result, the model is less complex and prevents overfitting. Moreover, the autoencoder is an unsupervised model and does not require labeled data.

In addition, during the experiment, there was 37% noise applied on the input images. However, the model could reconstruct the images to the original input data.

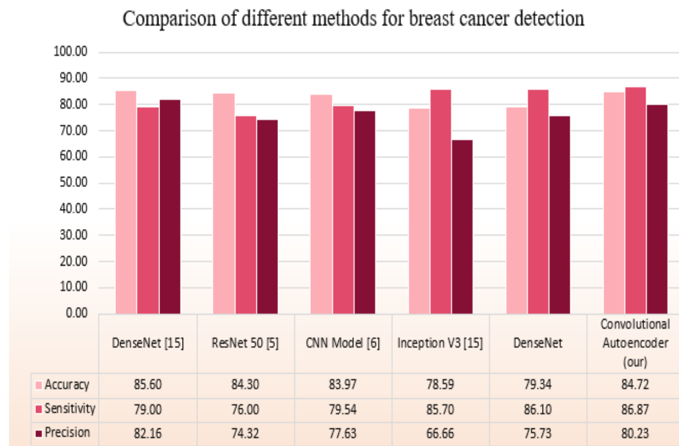

Fig. 7. Result of comparison different methods for detecting breast cancer.

References

[1] Polat, K. A new classification method for breast cancer diagnosis: feature selection artificial immune recognition system (FS-AIRS) / K. Polat, S. Sahan, H. Kodaz et al. // In: Advances in Natural Computation. – 2005. – pp. 830–838.

[2] Senapati, M. R. Local linear wavelet neural network for breast cancer recognition / M. R. Senapati, A. K. Mohanty, S. Dash et al. // Neural Computing and Applications. – 2013. – 22 (1). – 125–131.

[3] Marcano-Cedeno, A. Breast cancer classification applying the artificial metaplasticity algorithm / A. Marcano-Cedeno, J. Quintanilla-Dominguez, D. Andina // Neurocomputing. – 2011. – 74 (8). – 1243–1250.

[4] Diagnostic Mammogram,https://www.nationalbreastcancer.org/diagnostic-mammogram.

[5] Breast cancer: prevention and control,https://www.who.int/cancer/detection/breastcancer/en/.

[6] Inverse Relationship Between Precision and Recall,https://datascience.stackexchange.com/questions/49117/inverse-relationship-between-precision-and-recall.

[7] Arestaa G., Araujo T., Kwok S., Chennamsetty S. S., Safwan M., Alex V., Marami B., Prastawa M., Chan M., Donovan M., Fernandez G., Zeinehg J., Kohlh M., Walzi C., Ludwig F., Braunewell S., Baust M., Vu Q. D., Nhat To M. N., Kim E., Kwak J.T., Galal S., Sanchez-Freire V., Brancati N., Frucci M., Riccio D., Wang Y., Sun L., Ma K., Fang J., Kone I., Boulmane L., Campilho A., Eloy C., Polonia A., Aguiar P. BACH: grand challenge on breast cancer histology images. Medical Image Analysis. ELSEVIER 2019.

[8] BreakHist-Dataset-Image-Classification, https://github.com/Anki0909/BreakHist-Dataset-Image-Classification.

[9] Gao Huang, Zhuang Liu, Kilian Q. Weinberger. Densely Connected Convolutional Networks. arXiv:1608.06993v3. 2018.

[10] Majid Nawaz, Adel A. Sewissy, Taysir Hassan A. Soliman. Multi-Class Breast Cancer Classification using Deep Learning Convolutional Neural Network. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018.

[11] Yuexiang Li, Xinpeng Xie, Linlin Shen, Shaoxiong Liu. Reversed Active Learning based Atrous DenseNet for Pathological Image Classification. arXiv:1807.02420. 2018.

[12] Cruz-Roa, A., Gilmore, H., Basavanhally, A. et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep 7, 46450 (2017) doi:10.1038/srep46450.

[13] Sun, Wenqing Tseng, Bill Zhang, Jianying Qian, Wei. (2016). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Medical Imaging and Graphics. 57. 10.1016/j.compmedimag.2016.07.004.

[14] Zaychenko Y., Hamidov G. (2020) Hybrid Fuzzy CNN Network in the Problem of Medical Images Classification and Diagnostics. In: Liu Y., Wang L., Zhao L., Yu Z. (eds) Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery. ICNC-FSKD 2019. Advances in Intelligent Systems and Computing, vol 1074. Springer, Cham.

[15] Naderan M., Zaychenko Yu., Napoli A. Using convolutional neural networks for breast cancer diagnosing. System research and information technologies, 2019, № 4.