UDC 004.855.5
DOI: 10.20535/SRIT.2308-8893.2019.4.09

USING CONVOLUTIONAL NEURAL NETWORKS FOR BREAST CANCER DIAGNOSING

M. NADERAN, Yu. ZAYCHENKO, A. NAPOLI

Abstract. During the last few years, Convolutional Neural Networks (CNN) have been widely used in Computer-Aided Detection and the medical image analysis. The main idea of this paper is to modify CNN's architectures to achieve the better sensitivity and the precision for detecting breast cancer at an early stage compared to existing methods. For this purpose, several factors were considered before CNN training such as the data processing, model, dataset, etc. In the proposed model the following hyperparameters were the following: the dropout rate 0,2, epoch 38 and batch size 33. Besides the hyperparameters, two fully connected layers in the modified model were used. An average recall (sensitivity) in the recent works was 74%. The precision and recall of proposed model for breast cancer classification were 66,66% and 85,7%, respectively.

Keywords: convolutional neural networks, deep learning, computer-aided detection, breast cancer diagnosis, classification.

INTRODUCTION

Breast cancer is one of the most invasive cancers between women. The number of patients who have this type of cancer is increasing not only in poor countries but also in developed countries. Based on World Health Organization (WHO) [1], mammography is cost-effective for analyzing presence of breast cancer in patients. In other words, mammography is expensive but is the only method for screening breast cancer that has proven effective. In [2], the authors indicated that digital mammography (DM) produces the best medical image and is highly recommended for computer aided detection. In this paper, mammography scans were used for experiments, since they are of high quality.

The object of this study is mammography screening and the subject of this study is deep learning methods for diagnosing breast cancer.

In this paper, all methods for diagnosing breast cancer were implemented in jupyter notebook using keras libraries [3]. We consider recall and precision as evaluation metrics, since for detecting breast cancer, accuracy cannot be the only indicator for making decision.

NOVELTY

In this work, we proposed modified Inception V3 model to gain better results comparing to previous works. In order to that, the best values of hyperparameters that are playing an important role in the model were chosen. Also, before fully connected layer dropout with 0.2 rate was added for making model more independent to training data. Besides that, two fully connected layers were used in the modified model to achieve the goal.

RELATED WORKS

Systematic review (SR) has been done in [2]. In this paper authors illustrated how choosing the right medical images is important. Digital Mammography (DM) is the most important technique in medicine screening. DM helps to detect tumor before it develops further. Moreover, authors compare different methods that were used from 2011–2017 and conclude that SVM has the best result based on [4, 5, 6, 7].

Currently, there are few works that are considering using convolutional neural networks for the task. In [8], authors compare machine learning methods, such as Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k-Nearest Neighbors (k-NN) on white blood cell (WBC) datasets. Based on their experiments, SVM has the best accuracy with 97,13%. However, for cancer diagnosis task precision and recall should be considered as evaluation metrics. In [9], a modified CNN architecture with nine layers was proposed. From these layers, six of them were convolution and pooling layers and three were fully connected layers. The result of this work shows 69,99% and 81,44% accuracy for cancer classification and necrosis detection respectively.

Authors in [10] consider factors, such as true positive, false positive, precision and recall for classification using HPBCR (hybrid predictor of breast cancer recurrence). Sensitivity of the performance in [10] was 77% and the accuracy was 95%, and the same values for SVM and decision tree were 67%, 78% and 75%, 77% respectively.

Authors in [11] propose an [“end-to-end” approach in which a model to classify local image patches is pre-trained using a fully annotated dataset with region of interest (ROI)information].The experiment results show ResNet 50 has better accuracy (97%) comparing to VGG(84%).End-to-end learning process is a type of Deep learning process when all the parameters are trained jointly, rather than step by step. The weakness of [11] is that authors are using simple validation. Simple validation means that the training dataset is split in training and validation sets such as 70% of the data is used for training and 30% for validation. Instead K-fold cross validation divides the data into K number of sections/folds where each fold is used as a testing set at some point. Using K-fold cross validation helps to prevent overfitting without losing any data [12, 13].

CONVOLUTIONAL NEURAL NETWORK

A Neural Network (NN) is a network of neurons that are used to process information. A simple NN includes three layers: input, hidden and output. A Convolutional Neural Network (CNN) is a Deep Learning network which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. Convolutional neural network has three main layers: Convolutional layer, Pooling layer and Fully Connected layer. Fig. 1 illustrates the architecture of CNN. The main difference between Convolutional Neural Network CNN and Neural Network (NN), is a convolutional part. CNN has extra Convolutional and Pooling layers. Basically, a fully connected (FC) layer is simple NN. The number of convolutional and pooling layers depend on the model which is used. For example, in [9] there is no pooling layer but six convolutional layers.

There are three main operation illustrated in fig. 1:

1. Convolution and Non-Linearity (ReLU).
2. Pooling or Sub Sampling.
3. Classification (Fully Connected Layer).

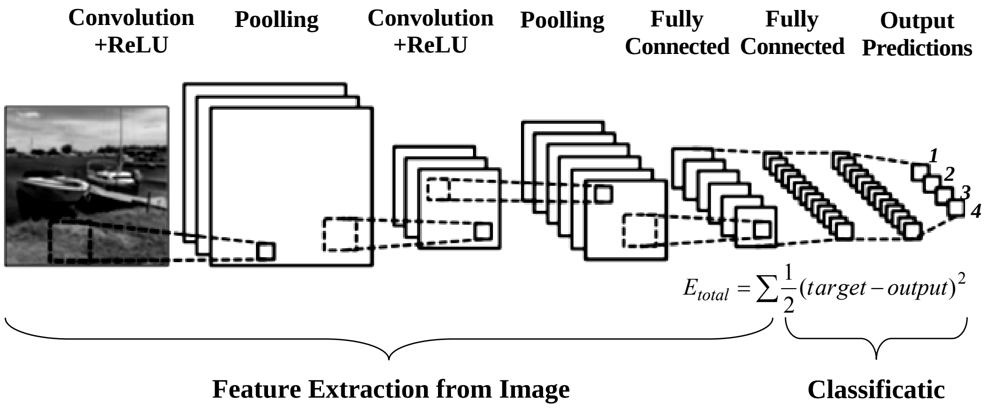

Fig. 1. A simple ConvNet; 1 — Dog (0), 2 —Cat (0), 3 — Boat (1), Bird — (0)

These operations are the main building blocks of each convolutional neural network. The first three operations are used for features extraction and the outputs of convolutional part (convolutional, ReLU and pooling) are used as input to the fully connected layers where classification happens.

The convolutional layer and ReLU function

The main idea of convolutional layer is features extraction. At this stage filters are applied to the input image for features extraction. In order to this, the filter slides (orange matrix is called "kernel") over the image (green matrix) by 1 pixel (stride) for every position, element wise multiplication is computed (between the two matrices) and outputs are added in order to get the final integer that forms a single element of the output matrix (pink matrix), (fig. 2). For example, the number "4", the first pixel on left side, calculated as $4 = 1*1 + 1*0 + 1*1 + 0*0 + 1*1 + 1*0 + 0*1 + 0*0 + 1*1$. By applying one filter to the image, we will get the first feature. By applying several filters to the same image, we will get several features that constitute a feature map [14].

The ReLU operation (also called activation function) is performed right after convolution. At this stage, all negative pixels are changed to zero in order to introduce non-linearity in ConvNet, since most of the real world data would be nonlinear (Convolution is a linear operation – element wise matrix multiplication and addition, so we account for non-linearity by introducing a non-linear function like ReLU) [23]. Operation ReLUis calculated using the formula below:

$$Y = \max(0, X).$$

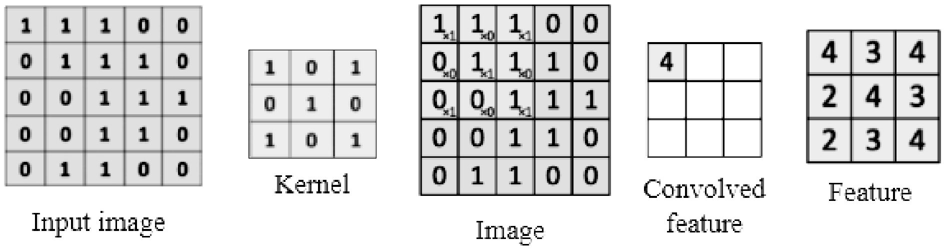

Fig. 2. Convolutional operation on one input image

The pooling layer

Pooling layer (also called subsampling), (fig. 3) changes the dimensionality of the feature maps. The input for this layer is the output of convolution (feature maps) and output is compressed version of the feature maps. Pooling can be calculated by Max, Average or Sum operation. The example of max pooling is illustrated in figure 2 where filter 2×2 with stripe 2 is applied to one feature. In this case, we slide filter over the feature map with stride equal 2. For instance, in order to calculate the first pixel of max pooling matrix in this example "6", the follow operation will be considered: $\text{Max}(1,1,5,6) = 6$

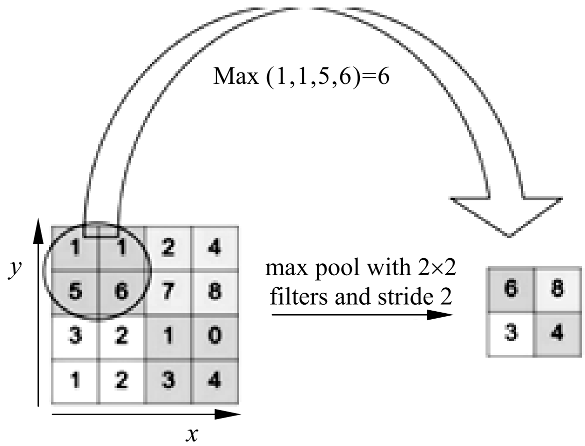

Fig. 3. Max pooling operation on one of the features

The fully connected layer

The Fully Connected layer is a traditional Multi-Layer Perceptron that uses Softmax activation function in the output layer. This function calculates the probabilities of each target class for the given input. The output of the convolutional and pooling layers represents high-level features of the input image. The purpose of the fully connected layer is to use these features for classifying the input image into various classes based on the training dataset.

METHODS

There are various machine learning algorithms and methods that could be used for diagnosing breast cancer with mammography, such as Support Vector Machine, Random Forest, K-Nearest Neighbor, etc. Even though support vector machines are widely used for different tasks, and have shown good results, convolutional neural networks are showing better performance among computer vision algorithms because of the ability to extract important features. After each convolutional operation, some features will be extracted and passed to the deeper layers, these features will become more specific and unique for the input image. For instance, for detecting a face in the picture, at first layer lines and curves of nose are extracted. In deeper layers, these lines and curves will be used as one feature (nose) of the face. By doing the same process, features like eyes, ears, lips etc. are extracted from the input image.

In [15] three different CNN architectures Cifar Net, AlexNet and GoogLeNet (Inception) were compared. It requires 5,37 GB of memory and 2h49m to apply GoogleNet. While for applying CifarNet, it takes 2,25 GB and 7m16s.

Some of the existing CNN models are very well realized. In this case to improve the precision and recall for our task, we can fine tune pre-trained models. Because, training model from scratch sometimes does not give the expected results. Since, training network from the scratch requires so many training datasets. Unfortunately, it is challenging to get access to the vast mammography screens. Transfer learning is a machine learning technique that helps reuse previous models for a running task and improve the used model architecture to reach higher accuracy and f-score.

EXPERIMENT

For conducting experiment Jupiter notebook was used. The machine where the program was run had an Intel Core i7 processor and NVIDIA GeForce driver.

In this paper two open-source datasets BreaKHis, Breast histology and Kaggle [16,17,18] were used during the experiment. In BreaKHis dataset, 1271 images were used for training and 70 images were used for testing, in Breast histology set, 200 images are used for training and 37 are used for testing. The dataset in Kaggle is numerical and 500 data were used for training and 70 data were used for testing.

Results show that small dataset could affect the accuracy and CNNs cannot be trained well leading to overfitting or underfitting. Chart 1 (fig. 4) shows accuracy comparison of Inception V3 for BreaKHis and Breast histology datasets.

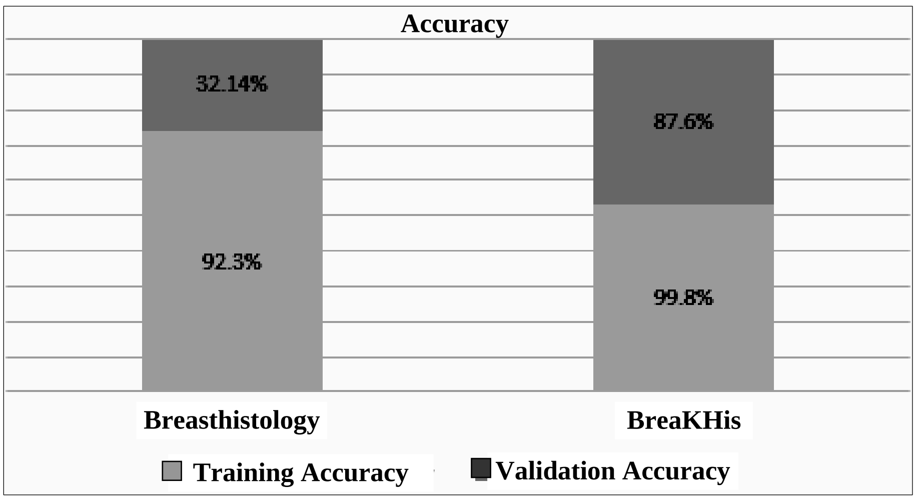

Fig. 4. Comparing accuracy for training and validation sets

According to the chart 1, accuracy for validation set in Breast histology dataset is noticeably lower value than in BreaKHis dataset. In this case our model is over fitted. This means that the model is doing well on the training set but not for those data which were never seen before (validation set). Thus, using a large dataset prevents model from overfitting during training and fine tuning.

Data preprocessing

Data preprocessing is an integral step in Machine Learning as the quality of data and the useful information that is derived from it directly affects the ability of the model. Also, it is used to transform the raw data in a useful and efficient format. Therefore, it is extremely important that the data are preprocessed before feeding it into the model [19]. Basically, we preprocess the raw data by importing libraries, read data, checking for missing values, checking for categorical data etc. Image Data Generator module [20] was used for augmentation data and make dataset bigger by creating different versions of one image.

After preprocessing data, we need to separate data into training and validation sets. Basically, validation set helps model to be well trained for data which were never seen before. There are two ways to separate the date into training and validation sets: simple validation and K-fold cross validation. Simple validation splits data into two sets where one part of data is used for training and another part is used for validation, whereas, in K-fold cross validation, dataset is split into K fold (part). At the first iteration, the first fold is used for validation set and (K–1) folds are used for training set. At the second iteration, the second fold is used for validation and the rest of the folds are used for training. This process is repeated until each fold has been used as validation set.

In Table 1, sensitivity (Recall) with K-fold validation is higher than with simple validation, 70% and 31,3% respectively. When simple validation is used during the experiment, some data could be missed in case when some images are not considered. As a result, CNN won't be trained well or over fitted.

Table 1. Comparing simple validation and K-fold validation for Inception V3

Simple validation			**K-fold validation**		
Precision	Recall	F1-score	Precision	Recall	F1-score
30,59%	31,3%	30,11%	65,7%	70%	67,78%

Table 2 shows the difference between our modified model and pre-trained model. Modified model is using the best weight to achieve better results, and based on Table 3, precision and recall (sensitivity) in the modified model are 66,66% and 85,7% respectively. In cancer diagnosing, the more important factor is sensitivity, which can be calculated as (6,1). In other word, we should avoid misclassification in case when the actual class is yes (cancer) and model prediction is no (no cancer).

Table 2. Comparing fine tune and pre-trained model (Inception V3)

Model	**Accuracy, %**	**Precision, %**	**Recall, %**	**F1-Score, %**
Inception V3	79,31	65,7	70	67,78
Modified Model	78,59	66,66	85,7	74,99

Based on Table 3, sensitivity (recall), precision and f1-score are calculated as following:

$$Sensitivity = \frac{TP}{Actual\ \ yes} = \frac{12}{14} = 85{,}7\%\,;$$

$$Precision = \frac{TP}{Predicted\ \ yes} = \frac{12}{18} = 66{,}66\%\,;$$

$$Sensitivity = \frac{Sensitivity * Precision}{Sensitivity + Precision} = 745{,}99\%\,.$$

T a b l e 3 . Confusion matrix

Index	Predicted No	Predicted Yes
Actual No	TN = 9	FP = 5
Actual Yes	FN = 1	TP = 12

Table 4 shows the comparison of machine learning algorithms on Kaggle dataset.

T a b l e 4 . Comparing machine learning algorithms on Kaggle dataset

ML algorithms	Accuracy	F1_score	Recall	Precision
LR	0,964912281	0,953846154	0,98412698	0,925373134
KNN	0,947368421	0,929133858	0,93650794	0,921875
SVM	0,959064327	0,945736434	0,96825397	0,924242424
NB	0,923976608	0,897637795	0,9047619	0,890625
DT	0,935672515	0,916030534	0,95238095	0,882352941
RF	0,964912281	0,950819672	0,92063492	0,983050847

Based on results in Table 4, Logistic Regression (LR) has shown better results comparing to other methods. Using LR method, we have achieved F1_Score and Recall 95,38% and 98,41% respectively. In future work, we will use CNN in numerical dataset and then will compare the obtained results with LR.

CONCLUSION AND DISCUSSION

Convolutional neural networks (CNN) is a special architecture of artificial neural networks. One of the most popular uses of this architecture is image classification. The reason is that it starts from lower abstraction and go deeper into higher abstraction. Convolutional neural network first starts to define curve and line in an image at higher layers and extracts general feature maps. While going deeper in lower (deeper) layers, the feature maps become more and more specific and unique.

The modified model has dropout rate = 0,2, epochs = 38 and batch size = 33. Besides choosing the right hyperparameters, in the model two fully connected layers were used. The model shows better result comparing the pre-trained model (Inception V3) with sensitivity of 85,7%.

Besides choosing the right model, the quality of the image is selected has an essential role on the result. In this paper, Digital Mammography (DM) is used, since the quality of scans in mammography is better than other medical scans. Even though DM is expensive but is the only effective method for screening breast cancer that has proven effective. Moreover, it helps to detect tumors at a very early stage.

In future, we plan to improve the architecture of the current model to reach better results.

REFERENCE

1. *Breast* cancer: prevention and control. — Available at: https://www.who.int/cancer/detection/breastcancer/en/
2. *Yassin N.* Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review / N. Yassin, S. Omran, E. El Houby, H. Allam // Computer Methods and Programs in Biomedicine. — 2018. — Vol. 156. — P. 25–45.
3. *Keras:* The Python Deep Learning library. — Available at: https://keras.io/
4. *Sharma S.* Computer-aided diagnosis of malignant mammograms using zernike moments and SVM / S. Sharma, P. Khanna // Journal of Digital Imaging. — 2015. — **28** (1). — P. 77–90.
5. *Beheshti S.M.A.* An efficient fractal method for detection and diagnosis of breast masses in mammograms / S.M.A. Beheshti et al. // Journal of Digital Imaging. — 2014. — **27** (5). — P. 661–669.
6. *Azar A.T.* Decision tree classifiers for automated medical diagnosis / A.T. Azar, S.M. El-Metwally // Neural Comput. Appl. — 2013. — **23** (7). — P. 2387–2403.
7. *Jian W.* Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform / W. Jian, X. Sun, S. Luo // Biomed. Eng. Online. — 2012. — **11** (1). — p. 96.
8. *Hiba A.* Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis / A. Hiba, M. Hajar, M. Hassan, N. Thomas // The 6th International Symposium on Frontiers in Ambient and Mobile Systems. Procedia Computer Science 83. — 2016. — P. 1064–1069.
9. *Sharmaa H.* Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology / H. Sharmaa, N. Zerbe, I. Klempert et al. // Computerized Medical Imaging and Graphics.— 2017. — **61.** — P. 2–13.
10. *Mohammad M.* A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning / M. Mohammad, M. Hamid, M. Marjan et al. // Computational and Structural Biotechnology Journal. — 2017. — **15.** — P. 75–85.
11. *Shen Li.* Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography / Li Shen, Laurie R. Margolies, Joseph H. Rothstein et al. Published online Mar 15. — doi: 10.1038/s41598-018-22437-z. 2018.
12. *K-Fold* Cross Validation. — Available at: https://medium.com/ datadriveninvestor/K-fold -cross-validation-6b8518070833
13. *Rakhlin A.* Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis / A. Rakhlin, A. Shvets, V. Iglovikov. — ICIAR 2018 Grand Challenge. arXiv:1802.00752v2. 2018.

14. *An Intuitive* Explanation of Convolutional Neural Networks. — Available at: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/
15. *Shin Hoo-Chang.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning / Hoo-Chang Shin, Holger R. Roth, Mingchen Gao et al. // IEEE Transactions on medical imaging. — 2016. — Vol. 35, N 5.
16. *Breast* cancer dataset from breakhis. — Available at: https://www.kaggle.com/kritika397/breast-cancer-dataset-from-breakhis
17. *Bioimaging* Challenge 2015 Breast Histology Dataset. — Available at: https://rdm.inesctec.pt/dataset/nis-2017-003
18. *Breast* histopathology. — Available at: https://www.kaggle.com/ paultimothy-mooney/breast-histopathology-images#IDC_regular_ps50_idx5.zip
19. *Sharmaa H.* Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology / H. Sharmaa, N. Zerbe, I. Klempert // Computerized Medical Imaging and Graphics. — 2017. — **61**. — P. 2–13.
20. *Image* Preprocessing. — Available at: https://keras.io/preprocessing/image/

Received 21.11.2019

From the Editorial Board: the article corresponds completely to submitted manuscript.