

Using Wasserstein Generative adversarial network For generating Artificial Medical Images and Convolutional Neural Network For Skin Cancer Detection

1st Pooyan Sedigh

*Department of Biomedical Engineering,
Science and Research Branch, Islamic Azad University,*
Tehran, Iran.

sedigh.pooyan@taarlab.com

Rasoul Sadeghian

Human and Robot Interaction Laboratory,
Tehran, Iran.

rasoulsadeghian1987@gmail.com

Shahrooz Shahin

Human and Robot Interaction Laboratory,
Tehran, Iran.

shahrooz.shahin@taarlab.com

Mehdi Tale Masouleh

*Human and Robot Interaction Laboratory, School of Electrical
and Computer Engineering, University of Tehran,*
Tehran, Iran.

m.t.masouleh@ut.ac.ir

Abstract—One of the big concern in skin cancer, is to detect the cancer with the high accuracy and so fast. This paper presents the process of detecting the skin cancer by using the Convolutional Neural Networks (CNN). Because of the lack of dataset in order to use for training the CNN algorithm, in the first step the Wasserstein Generative adversarial network (WGAN) algorithm has been used to create artificial members for the under study dataset. According to the results of implementing the designed CNN model on the proposed dataset, 95 percentages of the skin cancer images are detected correctly.

Index Terms—Skin cancer detection, Convolutional Neural Networks, Wasserstein Generative adversarial network, Artificial Intelligence, Neural Network

I. INTRODUCTION

Cancer can be regarded as one of the main reasons of human mortality. Among many different kinds of cancers, the one which is known to be more widespread between women is breast cancer [1]. From a recent statistics; it reveals that 6.3 million women suffered breast cancer and have cancer since five years before 2012. Moreover, in 2012, 1.7 women who have breast cancer were diagnosed [2]. The fast and high accurate diagnoses about cancers can reduce the risk of death among the patients [3]. According to the proposed demand, using the artificial intelligence method based on computer aid can consider as a perfect way [4]. In recent two decades, the significant progress in machine learning algorithms and the computers' processing power have been caused a significant growth in using the proposed algorithms in different fields like medical engineering [5]–[7]. One of the main applications of machine learning algorithms in field of medical engineering is cancer detection [8]–[10]. Based on the proposed application, the data as a data set are imported to the algorithm. There are

some dataset, which are used for breast cancer like, Wisconsin breast cancer [11].

Some of algorithms which have been used for the cancer detection can be regarded as, logistic regression, decision tree and Support Vector Machine (SVM) [12], [13]. By using the machine learning algorithms cancer diagnose, the cancer detection speed and quality have been increased and consequently the cancer detection's expenses may reduce, significantly.

The main contributions of this paper consist in, representing three machine learning algorithms, K-nearest neighbours, decision tree and logistic regression for cancer detection. Then implementing the proposed algorithms on Wisconsin breast cancer dataset. Moreover, comparing the aforementioned algorithms in terms of breast cancer detecting accuracy rate and computing time.

The rest of this paper is organized as follows. In Section 2, the process of creating artificial medical image with WGAN are completely presented. Then, the CNN algorithm structure which is used to create a deep learning model for detecting the skin cancer are described. Finally, the paper concludes with some hints and remaining as ongoing works.

A. Dataset

B. Wasserstein Generative adversarial network

In the first trace, the Generative adversarial network (GAN) algorithm is implemented on the preliminary database in order to create the new data, to increase the number of database's members to increase the accuracy of CNN model. The International Skin Imaging Collaboration database has some limitation for its users to download the data from it.

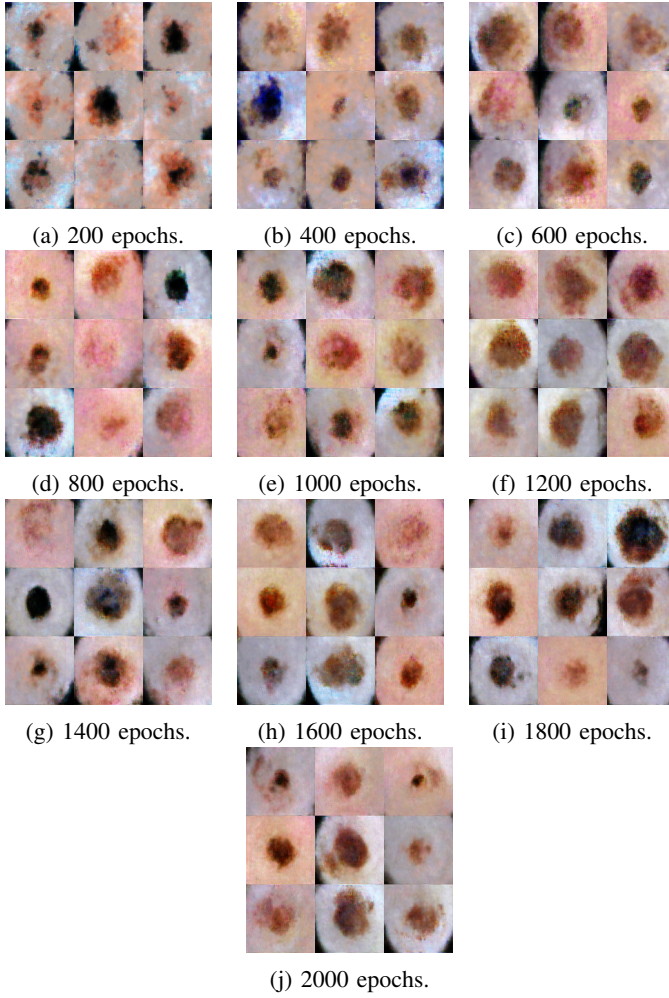


Fig. 1: WGAN performance during 2000 epochs.

[H] Under training data (s) $i=1$ to n Input a batch of k samples from random input, $\{z^1, \dots, z^m\}$;
 Input a batch of k data from generating distribution, $\{x^1, \dots, x^m\}$;
 Update the discriminator, Eq.?? Input a batch of k samples from random input, $\{z^1, \dots, z^m\}$;
 Update the generator, Eq.??

Based on the proposed limitation in the first step just 80 images are downloaded. Moreover, the downloaded images are divided in two parts as 55 and 25 for training and test part. The first part, training, is considered to use for generating artificial images from it. Figure ?? illustrates some of the 55 images, which are used as a database to use in GAN algorithm. In the following the progress of generating artificial images based on GAN method are represented, Figs. ?? and ?. According to the Figs. ?? and ??, nine artificial images are generated based on 2000 epochs. Figure 3 depicts the 2000th epoch. Based on Fig. ??, the performance of the designed GAN algorithm in terms of generating the new medical images is acceptable.

Then, in order to achieve better result the Wasserstein GAN(WGAN) algorithm is implemented on the data. The

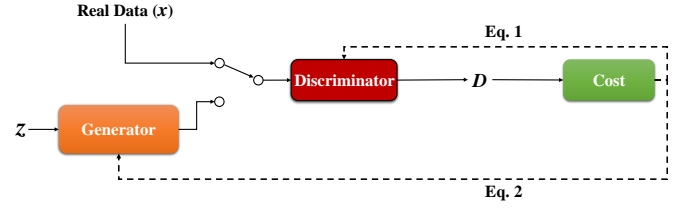


Fig. 2: The GAN algorithm structure.

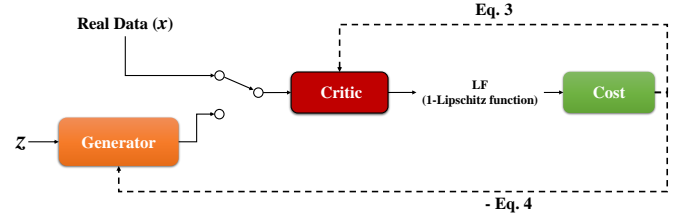


Fig. 3: The WGAN algorithm structure.

main difference between the GAN and WGAN can be regarded as, in the WGAN a new cost function using Wasserstein distance that has a smoother gradient everywhere is used, moreover, WGAN learns without paying attention to the generator act. Equations 1 and 2 represent the discriminator and generator functions, respectively, which are used for the GAN algorithm.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{n=1}^m [\log D(x^i) + \log(1 - D(G(z^i)))] \quad (1)$$

$$\nabla_{\theta_g} \frac{1}{m} \sum_{n=1}^m -\log(D(G(z^i))) \quad (2)$$

$$\nabla_{\omega} \frac{1}{m} \sum_{n=1}^m [f(x^i) - f(G(z^i))] \quad (3)$$

$$\nabla_{\omega} \frac{1}{m} \sum_{n=1}^m -f(G(z^i)) \quad (4)$$

Figures 2 and 3 illustrate the structure of GAN and WGAN algorithms. the variables which are used in Eqs. 1, 2, 3 and 4 are defined as, z is the noise which can consider as a normal or uniform distributions, G , is the generator and it used to create an image based on $x = G(z)$. Moreover, m is the number of noise and examples which are achieved from data generations, $D(x)$, is the function to calculate the probability that x came from the data rather than generator

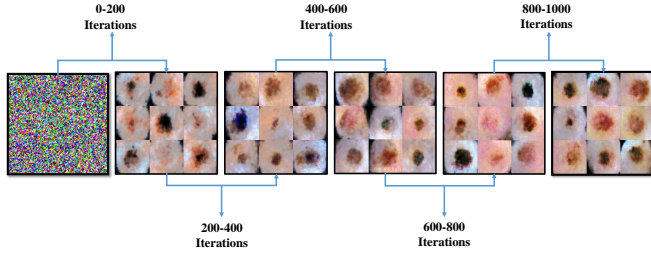


Fig. 4: The obtained result after 1000 iterations of WGAN implementation.

distribution. Furthermore, ∇ is the stochastic gradient decent which is used to training the GAN based on θ_d and θ_g parameters. Figure 4 depicts the obtained images after 1000 iterations. As it is obvious in the Fig. 4, after each 200 iterations the quality of the images have been improved.

C. Results

II. MODEL TRAINING

A. Convolutional Neural Network

B. Results

III. CONCLUSION

REFERENCES

- [1] Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F, "Cancer Incidence and Mortality Worldwide: IARC CancerBase," No 11. <http://globocan.iarc.fr>, 2013.
- [2] Shirazi AZ, Chabok SJ, Mohammadi Z., "A novel and reliable computational intelligence system for breast cancer detection," Medical & biological engineering & computing. 2018 May 1;56(5):721-32.
- [3] Khuriwal N, Mishra N. , "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm," In 2018 IEEMA Engineer Infinite Conference (eTechNxT) 2018 Mar 13 (pp. 1-5). IEEE.
- [4] Deng C, Perkowski M., "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer Detection," In 2015 IEEE International Symposium on Multiple-Valued Logic (ISMVL) 2015 May 1 (pp. 115-120). IEEE.
- [5] Qasem A, Abdullah SN, Sahran S, Wook TS, Hussain RI, Abdullah N, Ismail F., "Breast cancer mass localization based on machine learning," In Signal Processing & its Applications (CSPA), 2014 IEEE 10th International Colloquium on 2014 Mar 7 (pp. 31-36). IEEE.
- [6] Gayathri BM, Sumathi CP. , "Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer," In Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on 2016 Dec 15 (pp. 1-5). IEEE.
- [7] Forsyth AW, Barzilay R, Hughes KS, Lui D, Lorenz KA, Enzinger A, Tulskey JA, Lindvall C., "Machine learning methods to extract documentation of breast cancer symptoms from electronic health records," Journal of pain and symptom management. 2018 Jun 1;55(6):1492-9.
- [8] Nayak S, Gope D., "Comparison of supervised learning algorithms for RF-based breast cancer detection," In Computing and Electromagnetics International Workshop (CEM), 2017 2017 Jun 21 (pp. 13-14). IEEE.
- [9] Kim S, Jung S, Park Y, Lee J, Park J., "Effective liver cancer diagnosis method based on machine learning algorithm," In Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on 2014 Oct 14 (pp. 714-718). IEEE.
- [10] Turgut S, Datekin M, Ensari T., "Microarray breast cancer data classification using machine learning methods," In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) 2018 Apr 18 (pp. 1-3). IEEE.
- [11] Chaurasia V, Pal S, Tiwari BB., "Prediction of benign and malignant breast cancer using data mining techniques," Journal of Algorithms & Computational Technology. 2018 Jun;12(2):119-26.
- [12] Wang H, Zheng B, Yoon SW, Ko HS., "A support vector machine-based ensemble algorithm for breast cancer diagnosis," European Journal of Operational Research. 2018 Jun 1;267(2):687-99.
- [13] Zheng B, Yoon SW, Lam SS., "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications. 2014 Mar 1;41(4):1476-82.