شبكة المعلومات الجامعية

بسم الله الرحمن الرحيم

Ain Shams University Information Network

جامعة عين شمس

شبكة المعلومات الجامعية

@ ASUNET

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار

Scientific Computing Department
Faculty of Computer and Information Sciences
Ain Shams University

Computational Intelligence based Classification Method for Breast Cancer Diagnosis in Mammograms

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of PhD in Computer and Information Sciences
to
Department of Scientific Computing
Faculty of Computer and Information Sciences
Ain Shams University

By

Ghada Hamed Aly Kamel
Assistant Lecturer at Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Mohamed Fahmy Tolba
Professor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Assoc. Prof. Safaa Amin El-Sayed
Associate Professor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Assoc. Prof. Mohammed Abd El-Rahman Marey
Associate Professor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

August – 2021
Cairo

Acknowledgment

First of all, I would like to thank GOD for his endless blessings, for giving me the power and strength to complete this work and for giving me supportive people around me.

I would like to express my sincere gratitude to my supervisors; Prof. Dr. Mohamed Fahmy Tolba for his support (*god bless him*), patience and guidance, Dr. Safaa Amin El-Sayed for her continuous support and encouragement and Dr. Mohammed Marey for the special supervision experience he gave me. I am deeply thankful for being supervised by them.

I would like to thank my family especially my parents and sisters for being available all the time and for the love they gave me through the years. Thank you for accepting me through the tough times and for always believing in me. Especial heartfelt thanks to my husband for his deep support, endless encouragement, and his impressive patience and for sure being a main guide for me whenever I stuck or need him. Without all of them, it would have been much harder.

My dear friends who have helped me through the past time and kept on encouraging me to get this work done; Alaa Atef, Alaa Salah, Eman Hamdi, Hadeer El-Saadawy, Hanan Yousry, Samar Ahmed, Yasmin Mohamed and Yasmin Alaa.

Last but not least, I would like to thank all my professors, colleagues and students who kept on encouraging me. Thank you for being in my life.

Ghada

Abstract

With the recent research and development in deep learning since 2012, the emerging of Convolutional Neural Networks (CNNs) in bioinformatics, especially medical imaging, achieved obvious and tremendous improvement. Nowadays, from the raised critical challenges are the cancer detection in breast mammograms followed by the classification of the pathology of the localized lesions. Till now, the evaluation process of the screening mammograms is held by radiologists and physicians. Due to a large number of mammograms screened daily, this mammograms evaluation process becomes very monotonous, tiring, lengthy, costly, and significantly prone to errors. So, in the last two decades, the development of Computer Aided Detection (CAD) systems became very essential for early diagnosis. However, they should be improved for more accurate results which can help to obtain more confidence to the radiologist's decision.

In this thesis, I present the recently approaches that are machine learning based models developed to detect cancer in breast mammograms and classify them by analyzing them in the form of comparative study and analysis. Also, the mammographic datasets that are publicly available and popular as well are listed in the recent work to facilitate any new experiments and comparisons. It is conducted from the comparative study that the You Only Look Once (YOLO) model is one from the recent efficient and fast object detectors that obtains high accuracy compared with other detectors.

Based on the conducted comparative analysis, an end-to-end computer-aided diagnosis system based on You Only Look Once (YOLO) is proposed. The proposed system first preprocesses the mammograms from their DICOM format to images without losing data. Then, YOLO is utilized to take each mammogram and checks it in one shot to detect any existing lesions. Finally, the localized masses are classified into malignant and benign lesions without any human intervention.

YOLO has three different architectures, and in this thesis, the three versions are trained on breast mammograms for mass detection and classification to

compare their performance against each other. I utilized the anchor boxes upgrade in YOLO-V3 but in different manner. In order to achieve high detection accuracy, all anchor boxes used through training YOLO, are updated to sizes related to the masses I need to detect in mammograms, i.e., data related anchors. This is carried on by applying the k-means clustering on the sizes of all masses of the mammograms dataset to cluster them in a specific number of boxes which used later while YOLO training. Using the experimental results, it is proved how the idea of using YOLO-V3 by regenerated anchor boxes has a noticeable impact on the detection of masses and their classification as well. The proposed model is proved its ability to detect most of the challenging cases of masses and classify them correctly comparing with other recent detectors and the earlier versions of YOLO as well.

The existing publicly available datasets that contain fully field digital mammograms (FFDM) with lesions and at the same time contain accurate annotations, represented in only one dataset which is the INbreast. However, the disadvantage of the INbreast is its small number of samples. So, to utilize the good quality supported by the FFDM mammograms with handling the small number of samples issues, I implemented different augmentation approaches of other recent YOLO based studies. Among two commonly used techniques to augment data, it is proved by the results that augmenting the training set only is the fairest and accurate to be applied in the realistic scenarios.

The detected mammograms by the last proposed system have been checked to find out the problem of missing some masses that are already exist in mammograms. The main reason is that the masses exist in mammograms have no fixed sizes or near a specific range of sizes, they may be very small for example of width 10 and height 10 pixels and maybe large for example of width 900 and height 800 pixels. For this wide range I utilized YOLO-V3 model to detect masses through 3 phases-based architecture. First phase is the data preparation to convert DICOM files to images without losing data. Then, they are divided into mammograms with large and small masses representing the input to the second model training phase. The third phase is the model evaluation through two testing levels, first is the large masses checking and the second level is the small masses checking to output the intersected detection results for large and small masses. By this approach, the probability of missing large or small masses become very small. Moreover, this strategy has been proved its successfulness in overcoming the challenge of missing the small sized masses can be exist in breast.

Finally, YOLO is updated to a newer version which is YOLO-V4. In this thesis, the latest two models from YOLO (YOLOV3 and YOLO-V4) have been applied on mammograms. The main objective is to show how the new updates in YOLO-V4 will affect the prediction performance. Among the new updates in YOLO-V4 is the CSPDarknet53, which is a new Backbone that can enhance the learning capability of CNN. Besides the spatial pyramid pooling block which helps in increasing the receptive field and separating out the features of the most significant context. In YOLO-V3, Feature Pyramid Networks (FPN) is used for object detection while YOLO-V4 used PANet for different levels of detection. YOLO-V3 and YOLO-V4 are compared by evaluating their performance in detecting masses. It has been observed by the experimental results that YOLO-V4 outperformed YOLO-V3 by obtaining more accurate detection results and lower number of undetected mammograms, especially when its responsibility set to detect masses whatever their types instead of detecting benign and malignant masses.

To preserve the better accuracy obtained when YOLO is trained to only detect masses however its type, its classification role is replaced by other features extractors like ResNet and Inception. Inception-V3 results in better classification's performance than ResNet and near accuracy than YOLO's classifier. For this, I got confirmed from the experimental results that leaving the detection role to YOLO and then replace its features extractors by Inception-V3 to classify the localized objects is much more accurate and better than detect benign and malignant masses by YOLO, i.e., leaving the detection and classification roles to YOLO.

List of Publications

1- Hamed, Ghada, Mohammed Marey, Safaa El-Sayed Amin, and Mohamed F. Tolba. "A Proposed Model for Denoising Breast Mammogram Images." In 2018 13th International Conference on Computer Engineering and Systems (ICCES), pp. 652-657. IEEE, 2018.

2- Hamed, Ghada, Mohammed Abd El-Rahman Marey, Safaa El-Sayed Amin, and Mohamed Fahmy Tolba. "Deep Learning in Breast Cancer Detection and Classification." In Joint European-US Workshop on Applications of Invariance in Computer Vision, pp. 322-333. Springer, Cham, 2020.

3- Hamed, Ghada, Mohammed Abd El-Rahman Marey, Safaa El-Sayed Amin, and Mohamed Fahmy Tolba. "The Mass Size Effect on the Breast Cancer Detection Using 2-Levels of Evaluation." In International Conference on Advanced Intelligent Systems and Informatics, pp. 324-335. Springer, Cham, 2020.

4- Aly, Ghada Hamed, Mohammed Marey, Safaa Amin El-Sayed, and Mohamed Fahmy Tolba. "YOLO V3 and YOLO V4 For Masses Detection in Mammograms with ResNet and Inception for Masses Classification." In International conference on advanced Machine Learning Technologies and Applications, 2020.

5- Aly, Ghada Hamed, Mohammed Marey, Safaa Amin El-Sayed, and Mohamed Fahmy Tolba. "YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms." Computer Methods and Programs in Biomedicine (2020): 105823.

6- Aly, Ghada Hamed, Mohammed Marey, Safaa Amin El-Sayed, and Mohamed Fahmy Tolba. "Comparative Study and Analysis of Recent Computer Aided Diagnosis Systems for Masses Detection in Mammograms." International Journal of Intelligent Computing and Information Sciences 21, no. 1 (2021): 33-48.

7- Hamed, G., Marey, M., Amin, S. and Tolba, M.F., 2021. Automated Breast Cancer Detection and Classification in Full Field Digital Mammograms using Two Full and Cropped Detection Paths Approach. IEEE Access.

Table of Contents

List of Figures