بسم الله الرحمن الرحيم

MONA MAGHRABY

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

MONA MAGHRABY

شبكة المعلومات الجامعية

التوثيق الإلكتروني 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

MONA MAGHRABY

AIN SHAMS UNIVERSITY
FACULTY of COMPUTER &
INFORMATION SCIENCES
COMPUTER SCIENCE DEPARTMENT

Medical Image Classification based on Ensemble Machine learning Methods

A thesis submitted to the Department of Computer Science, Faculty of Computer and Information Science, Ain Shams University, in partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Science

By:

Nada Sherif Abdel Galil El Askary

Bachelor in Computer Science,
Faculty of Computer and Information Science,
Ain Shams University, Cairo, Egypt

Supervised By:

Prof. Dr. Mohamed Ismail Roushdy
Professor, Department of Computer Science
Faculty of Computers & Information Science
Ain Shams University

Dr. Mohammed A.-M. Salem
Associate Professor, Department of Scientific Computing
Faculty of Computer and Information Sciences
Ain Shams University
Faculty of Media Engineering and Technology
German University in Cairo

September 2020

Dedicated to my beloved father Sherif El Askary (Rahimahu Allah) who suffered alot from the cancer disease and passed away before seeing his daughter trying to find a solution to help him. . .

Declaration of Authorship

I, Nada Sherif ABDEL GALIL EL-ASKARY, declare that this thesis titled, 'Medical Images Classification Based on Ensemble Machine Learning Methods' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Nada Sherif Abdel Galil El-Askary

Date: May 2020

"*Who doesn't taste the bitter of Science 'education' an hour; he will sip the Humiliation of Ignorance all his life..*"

Imam Shafi'i (Rahimahu Allah)

"*Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.*"

Thomas A. Edison

"*Successful and unsuccessful people do not vary greatly in their abilities. They vary in their desires to reach their potential.*"

John Maxwell

AIN SHAMS UNIVERSITY

Abstract

Faculty of Computer and Information Science
Computer Science Department

Master Degree of Science Thesis

Medical Images Classification Based on Ensemble Machine Learning Methods

by Nada Sherif ABDEL GALIL EL-ASKary

Machine learning became a leading technology in all fields of life. In bioinformatics, different techniques are used to learn a machine how to detect and classify pathologies from digital medical images. These images could be of various types such as Computed Tomography (CT) and X-ray. As for lung CT images, machine learning can help radiologist to automatically detect lung nodules which can lead to lung cancer in their early stages and with high accuracy. Lung nodule is an abnormal growth in the lung. These nodules can be either benign (non-cancerous) or malignant (cancer) and in this later case the patient is suffering from pathological lung. Early detection of lung nodule decreases the risk of advanced stages in lung cancer disease and raises the possibility of saving precious human lives.
Random forest (RF), an ensemble machine learning algorithm, is used to detect the lung nodules and classify soft-tissues into nodules and non-nodules. A lung nodule classification approach is proposed to improve early detection for nodules in addition to optimizing RF and reached lung nodule localization. A five stages model has been built and tested using 214 cases from the LIDC database. Stage 1 is image acquisition and preprocessing. Stage 2 is extracting 119 features from each pixel in the lung CT image. Stage 3 is refining feature vectors by removing all duplicate instances and undersampling the non-nodule class. Stage 4 is tuning the RF parameters. Stage 5 is examining different collections from the extracted feature sets to select those scores best for classification. The accuracy achievedby RF is the highest compared to other machine learning classifiers such as KNN, SVM, and DT. The proposed method aimed to analyze and select features that maximize classification results. Pixel based feature set and wavelet based set scored best for higher accuracy. RF was tuned with 80 trees and 0.04 for in-bag-fraction. Best results were achieved by the proposed model are 94.57%, 98% and 96.28% for sensitivity, specificity, and accuracy respectively.

Acknowledgements

Thanks to God for leading me all the way though out my thesis journey.

Thanks to my supervisors; Prof. Dr. Mohamed Roushdy for guiding my to the right way and Dr. Mohammed Abd El Megeed for his endless knowleadge giving and advices for me to be more acheivable student. I would love also to thank Dr. Shereen Mousa for helping me in the publications processes.

Great thanks to my mother, my husband, my little children and all my family for supporting and being so kind to me while I was so depressed and frustrating. Their encouragement gave me the push to continue.

Finally, I should also acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health, and their critical role in the creation of the free publicly available LIDC/IDRI Database used in this study. . .

Contents

List of Figures

List of Tables

List of Publications and Submissions

List of published work

1. Nada S. El-Askary, Mohammed A.-M. Salem, and Mohamed I. Roushdy, "Lung Nodule Detection and Classification using Random Forest: A Review", In Proceedings of the 2019 9^{th} International Conference on Intelligent Computing and Information Systems (ICICIS '19), Cairo, Egypt, 2019, pp. 105-111, doi: 10.1109/ICICIS46948.2019.9014706.

2. Nada S. El-Askary, Mohammed A.-M. Salem, and Mohamed I. Roushdy. "Feature Extraction and Analysis for Lung Nodule Classification using Random Forest", In Proceedings of the 2019, 8^{th} International Conference on Software and Information Engineering (ICSIE '19), Cairo, Egypt, 2019. ACM pp. 248–252. DOI:https://doi.org/10.1145/3328833.3328872

List of submitted work to international journal

1. Nada S. El-Askary, Mohammed A.-M. Salem, and Mohamed I. Roushdy, "Features processing for Random Forest optimization in lung nodule localization" 2019