Computer Science Department
Faculty of Computer and Information Sciences
Ain Shams University

Diagnosis of Skin Cancer Using Machine Learning Techniques

Dissertation submitted to Computer Science Department
Faculty of Computer and Information Sciences, Ain Shams University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

By

Munya Abdulmajid Arasi

Lecturer Assistance of Information Technology Department
Faculty of Engineering, University of Aden, Yemen.

Supervised By

Prof. Dr. Abdel-Badeeh M. Salem
Professor of Computer Science
Faculty of Computer and
Information Sciences
Ain Shams University

Prof. Dr. El-Sayed M. El- Horbaty
Professor of Computer Science
Faculty of Computer and
Information Sciences
Ain Shams University

Prof. Dr. El-Sayed A. El-Dahshan
Professor at Faculty of Computer and Information Technology
The Egyptian University for E-learning

Cairo, 2018

Acknowledgement

First of all, I humbly give my deepest thanks to God for giving me the opportunity and the strength to accomplish this work.

I would like to thank all the people who helped me to make this work possible, especially my supervision committee. I would like to give my deep appreciation and thanks my supervisor, **Prof. Dr. Abdel-Badeeh M. Salem** for the confidence given to me and his wisdom intelligence, and creative view. His supervision inspired and helped me to complete the researches and writing this thesis, he always gave me the precise advice at the precise moment throughout my PhD study.

Also, I would like to express my sincere gratitude to my co-supervisor **Prof. Dr. El-Sayed M. El- Horbaty**, for always motivating me for going further and further. His enthusiasm, dedication and impressive depth of knowledge have been a great inspiration to me. I learned from him much about the field beyond my thesis topic and I know that this will help me throughout my career. He patiently helped me through my explorations and spared time to discuss ideas with me. I credit most of my intellectual growth during this period to him.

I am most grateful to my co-supervisor **Prof. Dr. El-Sayed A. El-Dahshan**, for his support and guidance of this study. He provided invaluable suggestion and encouragement from the beginning of my research. I benefited numerous discussions through his shared ideas, views from the perspective of their fields of expertise, and helped me when I needed some advice in different matters.

I also thank to the dean of the faculty of computer and information sciences, staff members, and registration committee for their advice and supporting.

Last but not least, I owe special thanks to my parents, my children Shahd and Abdullah and my sisters and brother for their patience with me, and their unlimited support and encouragement to do well.

Abstract

Skin cancer can be considered as one of the most growing types and deadly in the world of cancer; the importance of these tumors is the malignant melanoma. It is one the growing threats in the developed world, which could lead to increasing mortality. The main danger of melanoma is its ability to grow in any part in the body by lymphatic and blood vessels, making it difficult to diagnose in its early stages. As such, mortality rates increase by the time the melanoma is diagnosed in its late stages. Therefore, early diagnosis is crucial for reducing mortality and morbidity. Early stage diagnostic of benign and malignant melanoma is a crucial problem of skin cancer.

Nowadays, medical decision support employs computer technology, which helps the dermatologist to recognize a benign from malignant melanoma. Such technology takes less time, and gives accurate diagnosis with reduced related errors.

The thesis is divided into two parts; the first part summarizes the survey on machine learning techniques for diagnostic sciences in medical imaging, and focuses on a literature review of the Computer Aided Diagnosis (CAD) system that implemented in identifying and diagnosing malignant melanoma of dermoscopy images with emphasis on computer-aided image analysis starting with image acquisition and pre-processing, and finishing with the malignant melanoma classification of dermoscopy images. The techniques employed in each step are explained, and their benefits and limitations are identified. Apart from that, performance of a number of classifiers specifically developed for the diagnosis of malignant melanoma is compared.

The second part is based on various methodologies for CAD system, which are proposed on the problem of malignant melanoma detection using machine learning and deep learning techniques from macroscopic and dermoscopic images. These methodologies are composed of four stages: image acquisition, pre-processing, feature extraction, classification, and evaluation.

Different data sets are used like: Med-Node, Dermatology Information System (DermIS), and DermQuest. Various techniques are implemented in pre-processing such as: morphological operators, adaptive median filter and contrast adjustment. The three proposed techniques for feature extraction include Discrete Wavelet Transform (DWT), Principle Component Analysis (PCA), and the texture feature. These features serve as the input to the various classification techniques such as: K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), Neuro-Fuzzy, Decision Trees (DT), Naïve Bayes (NB), deep learning using Stacked Auto Encoders (SAEs). Different statistical tests are performed for analysis the macroscopic and dermoscopic images, and to verify and validate that the accuracy is true for comparing machine learning techniques. Also it is provided guarantee the reliability of study.

The experimental results show that the proposed methodologies are efficient, and achieving high classification accuracies in the range [75 % - 100 %] on macroscopic images, and achieving high classification accuracies in the range [90.0 % - 98.8 %] on dermoscopic images. We compared the obtained results with other techniques. The comparative results showed that the proposed CAD methodologies are more accurate than other techniques in this field of melanoma diagnosis.

Table of Contents

CHAPTER 4: Analysis of Different Techniques Used in CAD System for Melanoma Diagnostic

CHAPTER 5: Machine Learning Techniques For Macroscopic Images

CHAPTER 6: Machine Learning Techniques For Dermoscopy Images Based on Hybrid DWT and PCA Features

CHAPTER 7: Machine Learning Techniques For Dermoscopy Images Based on DWT and Texture Features

CHAPTER 8: Conclusions and Future Work

List of Tables

List of Figures

List of Abbreviations

ABCD	rule (A: Asymmetry, B: Boundary, C: Colour and D: Diameter of the lesion)
AC	Accuracy
AHE	Adaptive Histogram Equalization
ANN	Artificial Neural Network
ANOVA	Analysis Of Variance
AR	Association rules
BNs	Bayesian Networks
BPNs	Back propagation network
CAD	Computer Aided Diagnosis
CLAHE	Contrast Limited Adaptive Histogram Equalization technique
CNNs	Convolutional Neural Networks
CPN	Counter Propagation Network
CSLM	Confocal Scanning Laser Microscopy
CT	Computed tomography
DBNs	Deep Belief Networks
Deg-malig	Degree of malignancy
DLNN	Deep learning based Neural Networks
DR	Digital radiography
DSS	Decision support system
DT	Decision Tree
DWT	Discrete Wavelet Transform
EBPN	Error Back Propagation Network
FCM	Fuzzy C-means
FP-Tree	Frequent pattern tree
FTA	Fuzzy-T ARM
GLCM	Gray Level Co-occurrence Matrix
HNB	Hidden Naive Bayes
HSV	Hue-Saturation-Value
IHP	Independent Histogram Pursuit

kNN	k-nearest neighbor
LMT	Logistic Model Tree
LVQ	The Learning vector quantization
MLP	Multi-Layer Perceptron
MRI	Magnetic resonance images
MSIM	Multistage Illumination Modeling
NB	Naïve Bayes
NCI	National Cancer Institute
NFARM	Novel Fuzzy Association Rule Mining
NOS	Normalized Otsu's Segmentation
OCT	Optical Coherence Tomography
PCA	Principle Component Analysis
PNN	Probabilistic neural network
RBF	Radial Basis Function
RF	Random Forest
RNNs	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
SAEs	Stacked Auto Encoders
SE	Sensitivity
SLP	Single Layer Perceptron
SMO	Sequential Minimal Optimization
SOM	Self-Organizing Map
SP	Specificity
SVD	Singular Value Decomposition
SVM	Support Vector Machine
UMCG	University Medical Center Groningen
UV	Ultraviolet
WBCHIR	Wavelet Based Color Histogram Image Retrieval

List of Symbols

Symbol	Meaning
P(C)	prior probability
p(**x**|C)	class likelihood
P(C|**x**)	posterior probability
x	observation
x = (*x1,…, xn*)	a finite set of features
C	Confidence factor
x_i	sample
y_i	associated class label
(C_1, C_2 …. …., C_K)	The data is partitioned into *K* clusters
TPR	true positive Rate
TNR	true negative Rate
FPR	false positive Rate
FNR	false negative Rate
df	Degrees of freedom of the test
sd	standard deviation
ci	Confidence interval vector
H_0	Null Hypothesis
H_a	Alternate Hypothesis
µ	mean value
P-value	a probability –value
α	significance level