

(قَالَ يَا أَيُّهَا الْمَلَأُ أَيُّكُمْ يَأْتِينِي بِعَرْشِهَا قَبْلَ أَنْ يَأْتُونِي مُسْلِمِينَ ﴿٣٨﴾ قَالَ
عَفْرَيْتُ مِنَ الْجِنَّ أَنَا آتِيكُمْ بِهِ قَبْلَ أَنْ تَقُومَ مِنْ مَقَامِكُمْ وَإِنِّي عَلَيْهِ لَقَوِيٌّ
أَمِينٌ ﴿٣٩﴾ قَالَ الصَّادِقُ عَلَيْهِ عِلْمٌ مِنَ الْكِتَابِ أَنَا آتِيكُمْ بِهِ قَبْلَ أَنْ يَرْتَكِبَ
إِلَيْكُمْ حَارُفُكُمْ فَلَمَّا رَأَهُ مُسْتَقِرًّا عَلَيْهِ قَالَ لَهَا مِنَ فَضْلِ رَبِّي لِيَبْلُغَنِي
الْأَشْكُرُ أَمْ أَكْفُرُ وَمَنْ شَكَرَ فَإِنَّمَا يَشْكُرُ لِنَفْسِهِ وَمَنْ كَفَرَ فَإِنَّ رَبِّي

غَنِيٌّ كَرِيمٌ (النمل/ 38- 40

DEDICATION:

This thesis is dedicated to my mother and my father; it is also dedicated to my brothers , sisters , my colleagues , friends.

And to each of the lit up with his knowledge other's mind or solve the correct answer puzzled questioners he showed morals humility scientists and the beauty of those who know.

ACKNOWLEDGEMENT:

Praise to Allah, Lord of the Worlds, peace and prayers be upon human teacher Prophet Mahammed and his family and “sahaba” and who had followed them in a good way to the last day of the human's life.

My thanks to all who contributed in the output of this research, and to all those who were the cause of education and guiding and helping me.

And many thanks to supervisors Dr. Ali Ahmed who make full effort in guiding and steering while working in the research.

May Allah reward you and guide your steps.

ABSTRACT:

Breast cancer is one of the second leading causes of cancer death in women. Despite the fact that cancer is preventable and curable in primary stages, the huge number of patients are diagnosed with cancer very late. Conventional methods of detecting and diagnosing cancer mainly depend on skilled physicians, with the help of medical imaging, to detect certain symptoms that usually appear in the later stages of cancer. The objective of this study is to find the smallest subset of features and using ensemble method that can ensure highly accurate classification of breast cancer as either benign or malignant. In this study, ensemble classifier gives the maximum accuracy compared to individual classification classifiers.

المستخلص :

مرض سرطان الثدي من اكثر الامراض التى تودى لوفات العديد من النساء فى الأونة الاخيرة , على الرغم من أن مرض سرطان الثدي قابل للوقاية والعلاج فى المراحل الأولية. يتم تشخيص عدد كبير من المرضى المصابين بالسرطان فى وقت متأخر جدا. تعتمد الطرق التقليدية للكشف عن السرطان وتشخيصه بشكل رئيسي على الأطباء المهرة بمساعدة التصوير الطبي للكشف عن بعض الأعراض التى تظهر فى المراحل الأخيرة من المرض . الهدف من هذا البحث هو إيجاد تصنيف دقيق جدا لمرض سرطان الثدي . لتحقيق ذلك تم اختيار تقنية تنقيب البيانات كما تم اختيار منهجية التصنيف المتعدد لضمان جودة التصنيف كما تم تطبيق تقنية إختيار الخصائص باستخدام خوارزمية ال CFS لاختيار الخصائص المؤثرة فى التصنيف . أخيراً تمت مقارنة نتائج البحث مع المصنفات العادية مثل rep , j48 , randomforst . ووجدنا أن نتائج التصنيف المتعدد تعطى أعلى دقة من المصنفات الفردية .

Table of Contents:

الاية	i
DEDICATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
المستخلص	v
Table of Contents	vi
LIST OF FIGURES	ix
LIST OF Table	x
CHAPTER ONE	1
1. INTRODUCTION	2
1.1- Introduction	3
1.2- Research background	3
1.3- Problem statement	4
1.4- Objectives of the Study	4
1.5- Significant of the study	4
1.6- Scope of Study	4
1.7 Thesis Organization	5
CHAPTER TWO	
2.1- Introduction	6
2.2- FEATURE SELECTION METHODS	6
2.2.1 Information Gain	7
2.2.2 B. Relief	7
2.2.3 . Principal Component Analysis	8
2.2.4. Correlation Based Feature Selection	8
2.2.5. Consistency Based Subset Evaluation	9

2.3- General Approach of Classification	9
2.3.1. Decision Trees	10
2.3.2. Support Vector Machine	10
2.3.3. Genetic Algorithms Evolutionary Programming	10
2.3.4 Fuzzy Sets	11
2.3.5. Neural Networks	11
2.3.6 Ensemble Classifiers	11
2.3.6.1 Bagging	12
2.3.6.2 Boosting	12
2.4- Medical image classification	13
2.5 RELATED WORKS	14
CHAPTER THREE	
3.1- Introduction	15
3.2- Research phases	15
3.2.1 Phase1 Data Preprocessing	15
3.2.2 Phase2 classification	16
3.2.2.1 Ensemble classifier	16
3.2.2.2 Building classifier	17
3.2.2.3 Method of ensemble classifier	17
3.2.2.3.1 Bagging	17
3.2.2.3.2 Boosting	17
3.2.3 Phase 3 testing and evaluating	23
3.3 Correlation based Feature Selection(CFS) Algorithm	23
3.3.1 Material and Tools	25
3.3.2 Dataset	26
CHAPTER FOUR	
4.1 Introduction	28
4.2- Experiment and results	28

4.3 results	31
4.5 Discussions	37
4.4 Comparison with other study	37
CHAPTER FIVE	
5.1- Conclusions	40
5.2- Recommendation	40

LIST OF FIGURES:

Figure 3.1 Ensemble classifier	18
Figure 3.2 The components of CFS	26
Figure 3.2 Distribution of the patient based on the class label	29
Figure 4.1 Individual classification without data processing	29
Figure 4.2 Ensemble classification using j84,rep and random forest algorithms	30
Figure 4.3 Individual and ensemble classification accuracy without pre processing	31
Figure 4.4 Individual classification with pre processing	33
Figure 4.5 Ensemble classification with pre processing	34
Figure 4.6 The accuracies of ensemble methods and individual with training data set 65%, 70%, 80.	35
Figure 4.7 Evaluation of feature selection method and ensemble method	36

LIST OF TABEL

TABEL 3.1 Description of Breast Cancer Dataset	27
TABEL 3.2 Wisconsin Breast Cancer Dataset Attribute	28
TABEL 4.1 1 Individual and ensemble classification accuracies without pre processing	31
TABEL 4.2 Features obtain by features selection	32
TABEL 4.3 The accuracies of ensemble methods and individual with training data set 65%, 70%, 80.	35
TABLE4.4 Evaluation of feature selection method and ensemble method	31