

## ACKNOWLEDGEMENT

I would like to express my great appreciation to Dr. Shaza Mergani for her valuable suggestions during the planning and development of this research work.

I wish to thank also my parents, my husband and my cousin Dr. Safa for her help in collecting the data from patient medical reports.

## **Abstract**

Differential diagnosis means the process of differentiating between two or more diseases that share similar signs or symptoms, the differentiating in such cases consider as a challenge in the field of dermatology, a real Sudanese data was built for the purposes of this research which had been collected from the medical reports in Omdurman Military Hospital department of dermatology. Three models were built using three classification algorithms Naïve Bayes, j48 and IBK .The research aimed to build a model classifying four dermatology diseases which have high similarity in their symptoms, these diseases are:

- 1- Psoriasis
- 2 - seboreic dermatitis
3. lichen planus
- 4- cronic dermatitis

The classification models had an accuracy in the range of%90.6 to %99.4 ,the results showed that IBK algorithm gave the highest accuracy ( %99.4 ) and less time to construct the model.

## المستخلص

التشخيص التفريقي يعني عملية التفريق بين الأمراض ذات الأعراض المتشابهة ، ويعتبر التفريق في مثل هذه الحالات من التحديات التي تواجه الأطباء في مجال الأمراض الجلدية ، ولقد تم بناء بيانات سودانية حقيقية لأعراض هذا البحث تم جمعها من التقارير الطبية بمستشفى السلاح الطبي - أدرمان قسم الأمراض الجلدية ، تم بناء ثلاثة نماذج باستخدام ثلاثة خوارزميات تصنيفية هي Naïve Bayes ، j48 و IBK.

يهدف البحث إلى بناء نموذج يقوم بتصنيف أربعة أمراض جلدية تتشابه بشكل كبير في أعراضها ، وهذه الأمراض

هي:

1- الصدفية

2 - التهاب الجلد الدهني

3. الحزاز المسطح

4- التهاب الجلد المزمن

تراوحت الدقة التي تم الحصول عليها من نماذج التصنيف ما بين % 90.6 إلى % 99.4 ، وقد أظهرت النتائج أن

خوارزمية IBK أعطت أعلى دقة (% 99.4) وأقل وقت لبناء النموذج .

# Table of Contents

Chapter One : Introduction.....	1
1.1 Introduction .....	1
1.2 Problem statement .....	2
1.3 Research Significance .....	3
1.4 Hypotheses.....	3
1.5 Objectives.....	3
1.6 Research Scope .....	4
1.7 Research Organization .....	4
Chapter Two : Litruture review and related work .....	17
2.1 Introduction .....	5
2.2 Knowledge Discovery from Data (KDD).....	5
2.2.1 Data mining functionality.....	5
2.3 Data mining in medical domain.....	6
2.4 Related work .....	6
2.4.1 Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification .....	6
2.4.2 Classification of Dermatology Diseases through Bayes net and Best First Search.....	7
2.4.3 Data Mining in Dermatological Diagnosis: A Method for Severity Prediction .....	7
2.4.4 Diagnosing Common Skin Diseases using Soft Computing Techniques .....	7
Chapter Three : Methodology .....	8
3.1 Introduction .....	8
3.2 Weka Definition.....	8
3.3 Dataset Description .....	8
3.3.1 UCI Dermatology dataset description .....	8
3.3.2 Sudanese Dataset description .....	11
3.4 Data Preprocessing:.....	12
3.5 Classification and Prediction .....	13
3.5.1 Bayesian Classification .....	13
3.5.1.1 Bayesian Theorem .....	13
3.5.1.2Naïve Bayesian classification .....	14
3.5.2 Decision tree algorithm J48 .....	15
3.5.3 Lazy Classifier .....	15
3.5.3.1IBK Classifier(K-Nearst Neighbor) .....	16
Chapter Four : Results, Analysis and Discussion.....	17

4.1 Introduction: .....	17
4.2.1 First Experiment:.....	17
4.2.2 Second Experiment: .....	18
Chapter Five : Conclusion and Recommendation.....	21
5.1 Introduction: .....	21
5.2conclusion: .....	21
5.3 Recommendation:.....	21
References .....	23
Appendix.....	24-31

## List of Figures

Figure1. 1Differential Diagnosis.....	3
Figure 4. 1 Accuracy of the three algorithms when applied on UCI dataset.....	18
Figure 4. 2Compare the accuracy of the three models (second experiment ) .....	19

## List of Tables

Table 3. 1Class Distribution for the UCI dataset : .....	<b>Error! Bookmark not defined.</b>
Table 3. 2Class Distribution for the Sudanese dataset : .....	12
Table 4. 1Accuracy of the three algorithms when applied on UCI dataset.....	17
Table 4. 2confusion matrix when apply naïve byes on Sudanese dataset .....	<b>Error! Bookmark not defined.</b>
Table 4. 3confusion matrix when apply IBK on Sudanese dataset .....	19
Table 4. 4Accuracy measures and time for the three models (second experiment).....	20