



شبكة المعلومات الجامعية

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ





شبكة المعلومات الجامعية



شبكة المعلومات الجامعية

التوثيق الالكتروني والميكرو فيلم

جامعة عين شمس

التوثيق الالكتروني والميكرو فيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد اعدت دون أية تغيرات



يجب أن

تحفظ هذه الأفلام بعيداً عن الغبار

في درجة حرارة من 15 – 20 مئوية ورطوبة نسبية من 20-40 %

To be kept away from dust in dry cool place of
15 – 25c and relative humidity 20-40 %



شبكة المعلومات الجامعية



بعض الوثائق الأصلية تالفة



شبكة المعلومات الجامعية



بالرسالة صفحات
لم ترد بالأصل

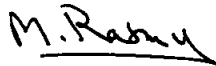
DATA MINING BASED ASSISTANT TOOL FOR EVALUATING PATIENTS OF CHRONIC LIVER DISEASES

By

Eng. Ahmed Mohamed Hashem Fathy Mahmoud

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirement for the Degree of
MASTER OF SCIENCE
In
SYSTEMS AND BIOMEDICAL ENGINEERING

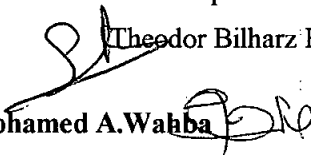
Under the Supervision of



Prof. Dr. Mohamed Emad M. Rasmy
Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

Prof. Dr. Sobhy Ahmed Kishta

Department of Surgery
Theodor Bilharz Research Institute


Dr. Khaled Mohamed A. Wahba

Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

March 2006

DATA MINING BASED ASSISTANT TOOL FOR EVALUATING PATIENTS OF CHRONIC LIVER DISEASES

By

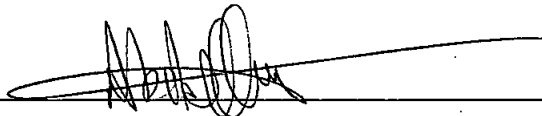
Eng. Ahmed Mohamed Hashem Fathy Mahmoud

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirement for the Degree of
MASTER OF SCIENCE
In
SYSTEMS AND BIOMEDICAL ENGINEERING

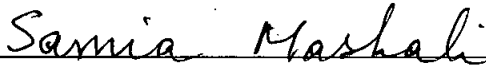
**Approved by the
Examining Committee**



Prof. Dr. Mohamed Emad M. Rasmy, Thesis Main Advisor



Prof. Dr. Abdallah Sayed Ahmed, Member



Prof. Dr. Samia Abdel Razek Mashali, Member

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

March 2006

CONTENTS

CONTENTS.....	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
ACKNOWLEDGEMENTS.....	xiii
ABSTRACT.....	xiv
CHAPTER 1: INTRODUCTION.....	1
1.1 OVERVIEW.....	1
1.2 PROBLEM DEFINITION.....	1
1.3 THESIS OBJECTIVES.....	2
1.4 THESIS METHODOLOGY.....	3
1.4.1 Methods used to obtain all the required information for the research.....	3
1.4.2 Methods to facilitate the understanding of data mining process.....	3
1.4.3 Methods used in the data mining process.....	4
1.5 THESIS STRUCTURE.....	4
CHAPTER 2: DATA MINING OVERVIEW.....	6
2.1 MOTIVATION OF DATA MINING.....	6
2.2 DATA MINING DEFINITION.....	6
2.3 ARCHITECTURE OF DATA MINING SYSTEM.....	8
2.4 DATA MINING FUNCTIONALITIES.....	10
2.4.1 Class Description: Characterization and Discrimination.....	10
2.4.2 Mining Frequent Patterns, Associations, and Correlations.....	10
2.4.3 Classification and Prediction.....	11
2.4.4 Cluster Analysis.....	11
2.5 POTENTIAL APPLICATION OF DATA MINING.....	12
2.6 DATA MINING PROCESS.....	13
2.6.1 The CRISP-DM Reference Model.....	13

2.7 DATA MINING TECHNIQUES.....	17
2.7.1 Statistical Techniques.....	17
a. Linear Regression.....	17
b. Logistic Regression.....	19
c. Factor Analysis / Principal Components Analysis.....	20
2.7.2 Classification Technique: Decision Tree Induction.....	22
2.7.3 Clustering Techniques.....	27
a. Two Step Clustering Technique.....	27
b. The K-Means Method.....	28
2.7.4 Association Rules Techniques.....	30
 CHAPTER 3: CHRONIC LIVER DISEASES.....	31
3.1 OVERVIEW.....	31
3.2 LIVER STRUCTURE.....	31
3.3 MAIN LIVER FUNCTIONS.....	33
3.4 CHRONIC LIVER DISEASES.....	37
3.5 PROGNOSIS OF CHRONIC LIVER DISEASES.....	40
3.5.1 Historical Background of Prognostic Models.....	41
3.5.2 MELD vs. Child-Pugh scores.....	50
3.5.3 MELD vs. Child-Pugh and Modified Child-Pugh scores.....	53
3.6 CONCLUSION.....	56
 CHAPTER 4: CRISP-DM IMPLEMENTATION.....	57
4.1 OVERVIEW.....	57
4.2 BUSINESS UNDERSTANDING.....	58
4.2.1 Determine the Business Objective.....	59
4.2.2 Assessing the Situation.....	59
4.2.3 Determine the Data Mining Goal	60
4.3 DATA UNDERSTANDING.....	61
4.3.1 Defining Data Resources.....	63
4.3.2 Data Base Construction.....	70

4.4 DATA PREPERATION.....	75
4.4.1 Overview.....	75
4.4.2 Data Selection.....	76
4.4.3 Data Cleaning.....	77
4.4.4 Data Integration.....	83
4.4.5 Data Formatting.....	83
4.5 CONCLUSION	83
CHAPTER 5: ROUGH FEATURE EXTRACTION.....	84
5.1 OVERVIEW.....	84
5.2 STATISTICAL MEASURES OF DATA.....	84
5.2.1 Patient Baseline Characteristics.....	84
5.2.2 Descriptive Statistics for Quantitative Variables.....	86
5.3 GRAPHICAL REPRESENTATION OF DATA.....	89
5.3.1 Distribution of Categorical Parameters of Child-Pugh classification.....	90
5.3.2 Histogram for Numeric Parameters of Child-Pugh classification.....	92
5.3.3 Exploratory Graphs of Child-Pugh Classification Parameters.....	103
5.4 CONCLUSION.....	111
CHAPTER 6: MODELLING, EVALUATION, AND DEPLOYMENT.....	113
6.1 OVERVIEW.....	113
6.2 MODELLING.....	113
6.2.1 Modeling Technique Selection.....	114
6.2.2 Generating Test Design.....	115
6.2.3 Model Building.....	116
6.2.4 Model Assessment.....	117
6.3 MODELLING USIN C5.0	118
6.3.1 Modeling Technique Selection.....	118
6.3.2 Model Building.....	118
6.3.3 Model Assessment.....	131

6.4 MODELLING USIN K-MEANS.....	142
6.4.1 Modeling Technique Selection.....	142
6.4.2 Model Building.....	142
6.4.3 Model Assessment.....	152
6.5 EVALUATION.....	159
6.5.1 Results Evaluation.....	159
6.5.2 Process Review.....	160
6.5.3 Determination of the Next Steps.....	161
6.6 DEPLOYMENT.....	161
6.6.1 Plan Deployment.....	161
6.6.2 Plan Monitoring and Maintenance.....	162
6.6.3 Producing Final Report.....	162
6.6.4 Project Review.....	162
CHAPTER 7: CONCLUSION.....	164
7.1 CONCLUSION.....	164
7.2 SUGGESTIONS FOR FUTURE WORK.....	165
REFERENCES.....	166
Appendix A: Description of Laboratory Tests.....	175
Appendix B: Data Collected Form for Cirrhotic Patient	181

LIST OF TABLES

TABLE 3.1	Routinely available tests of liver functions.....	36
TABLE 3.2	General investigations in the patient with cirrhosis.....	39
TABLE 3.3	The Child–Turcotte classification.....	42
TABLE 3.4	The Child-Turcotte-Pugh score.....	42
TABLE 3.5	Prognostic variables in primary biliary cirrhosis.....	46
TABLE 3.6	Mayo natural history model for primary biliary cirrhosis.....	47
TABLE 3.7	Comparison of Child-Pugh and MELD scores.....	50
TABLE 3.8	Comparison of the prognostic accuracy of Child-Pugh and MELD scores measured by concordance (c) statistic in several series.....	51
TABLE 3.9	Comparison of prognostic accuracy of MELD, CP and Modified CP scores for 3-month survival in patients with chronic liver disease.....	53
TABLE 3.10	Advantages and disadvantages of some prognostic models.....	55
TABLE 4.1	The main tasks that performed in Data Understanding phase.....	62
TABLE 4.2	Advantages and disadvantages of alternative study designs.....	65
TABLE 4.3	The parameters of inclusion criteria for cirrhotic patients.....	71
TABLE 5.1	Patient baseline characteristics.....	84
TABLE 5.2	Descriptive statistics for quantitative variables.....	87
TABLE 6.1	Major changes in the accuracy due to changes in C5.0 parameters, expert options, and simple options.....	122
TABLE 6.2	Major changes in the accuracy due to changes in C5.0 parameters using Cross Validation.....	124
TABLE 6.3	Sensitivity and specificity measures of the generated rules.....	141
TABLE 6.4	Comparison between accuracy of generated K-Means models using Total Complete Dataset vs. Equal Sampling Dataset.....	149

